The role of nuclear protein HBP1 in differentiation and proliferation using conditionally deficient mice.

Jasper de Boer

Research Thesis submitted for the degree of doctor of philosophy of the

University of London

Supervisor: Dr Dimitris Kioussis

Division of Molecular Immunology

National Institute for Medical Research

Mill Hill, London

2004

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

HBP1 (HMG box containing protein 1) is a member of the high mobility group chromosomal proteins. The HMG box contained in this protein binds DNA in a sequence specific manner and other domains in the protein allow its participation in protein-protein interactions that can have profound effects in the regulation of chromatin structure and, as a consequence, gene expression, proliferation and differentiation programmes. The protein also contains a novel domain called AXH that is also present in Ataxin-1 (ATX1), a human protein responsible for spinocerebellar ataxia type 1. The function of the AXH domain is unknown.

An initial study to define the function of the AXH domain was to measure its structural properties and stability. This was achieved by producing protein constructs spanning the AXH modules of ATX1 and HBP1 and comparing their properties. This led to the identification of the minimal region sufficient for forming independently folded units (domains). It was also shown that the AXH of ATX1 contains a dimerization domain whereas the AXH of HBP1 stays as a monomer.

In order to define the subcellular localisation of HBP1 the protein was tagged with enhanced yellow fluorescent protein (EYFP) and expressed in HEK293 cells. Subcellular localization of the respective fusion proteins was analyzed by direct fluorescence microscopy. EYFP-tagged wild type HBP1 was located exclusively in nuclei, while EYFP alone was equally distributed throughout the cells. To identify the NLS, several HBP1 deletion mutants were tagged with EYFP. Mutant HBP1 that contained only the HMG box also localized exclusively in nuclei, while the deletion

of the HMG box resulted in the loss of preference for nucleus. These results indicate that HBP1 is a nuclear protein and that the NLS is located within the HMG box of HBP1.

To assess the function of HBP1, standard gene-targeting techniques were used to produce a mouse in which the exons coding for the HMG box, an essential region of HBP1, is flanked by loxP sites (floxed), so that this domain can be deleted specifically at will by directing the expression of Cre in a tissue specific manner.

Using the above strategy two mouse lines were generated: one carrying HBP1alleles with floxed HMG box coding exons to be used in conditional deletion studies and one with a constitutive deletion of these exons (ΔHMG). Furthermore, we generated two Cre transgenic mice lines that allow the deletion of floxed gene segments either only in the lymphoid compartment (hCD2-iCre) or in the entire haematopoietic compartment (Vav-iCre). Crossing these Cre transgenic lines to mice carrying the floxed alleles will allow the inactivation of genes only in certain tissues so that biological questions can be answered with increased specificity.

Mice homozygous for the gene deletion ($HBP1^{\Delta HMG/\Delta HMG}$ mice) were viable, fertile, and have normal numbers and percentages of haematopoietic lineages. However, T lymphocytes from $HBP1^{\Delta HMG/\Delta HMG}$ animals proliferate at a significantly higher rate in response to anti-CD3 in vitro. Initial experiments indicate that the increased proliferation by $HBP1^{\Delta HMG/\Delta HMG}$ T lymphocytes might be due to the fact that the cells are in a "poised for activation" state and, thus, predisposed for immediate response to stimulation.

Acknowledgements

This thesis could not have been written without the assistance, help, encouragement and support of many people. The foremost among them is my supervisor, Dimitris Kioussis. If not for his advice, commitment, and inspiration this work would not be to the standard it is now.

I have to give credit to all members of our laboratory and division, who helped through the last couple of years. I am especially grateful to Mauro Tolaini for generating transgenic mice, Adam Williams for his enthusiastic and friendly collaboration in establishing and analysing the iCre transgenic mouse lines. I would like to thank Eleni Ktistaki for the help with the transfection experiments, Anna Garefalaki and Ursula Menzel for the help and advice on the ES cell work and finally, I would like to thank Cesira de Chiara and Annalisa Pastore for the collaboration on studying the AXH module. My special thanks go to Trisha Norton and Keith Williams for the excellent maintenance of the mice.

It almost goes without saying that my thesis could not have been completed without the support of my wife Kym and my parents. It is because of their love and support that I've been able to complete the requirements for my degree by writing this thesis.

Jasper de Boer

January 2004

Contents

List of Figures

Abbreviations

2-ME	2-mercaptoethanol
aa	amino acid
AMP	Ampicillin Resistance Marker
APC	Allophycocyanin
APS	Ammonium Persulfate
ATP	Adenosine Triphosphate
AXH	130 aa long motif identified in HBP1 and ATX1
bp	base pair
BSA	Bovine Serum Albumin
CD	Cluster of Differentiation
CDK	Cyclin Dependent Kinase
CDKI	CDK Inhibitor
cDNA	complementary DNA
CLP	Common Lymphoid Progenitor
CMP	Common Myeloid Progenitor
cpm	counts per minute
dCTP	deoxycytidine triphosphate
ddH2O	double distilled water
DAPI	4',6-Diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle's Medium

DMSO	Dimethylsulphoxide
DN	CD4-CD8- Double Negative
DNA	Deoxyribonucleic Acids
DNase 1	Deoxyribonuclease 1
DNMT	DNA Methyltransferases
dNTP	Deoxyribonucleotides
DP	CD4+CD8+ Double Positive
Dsh	Dishevelled
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetracetic Acid
EMSA	Electro Mobility Shift Assay
ERK	Extracellular signal Regulated Kinase
ES	Embryonic Stem cells
EYFP	Enhanced Yellow Fluorescent Protein
FACS	Fluorescence Activated Cell Sorting
FCS	Foetal Calf Serum
FITC	Fluoresceine isothiocyante
FT1	Footprint within hCD2 that binds HBP1
Fzd	Frizzled
GFP	Green Fluorescent Protein
GMP	Granulocyte Monocyte Progenitor
GTP	Guanosine Triphosphate
HAT	Histone Acetyl Transferase
HBP1	HMG Box containing Protein 1
HBS	Hepes Buffered Saline

hCD2	Human CD2 gene
HDAC	Histone Deacetylases
HEK 293	Human Embryonic Kidney cell line 293
HEPES	N-2 Hydroxyethylpiperazine-N-2-ethanol
HMG	High Mobility Group
HP1	Heterochromatin Protein 1
HS	DNase I hypersensitivity
HSC	Haematopoietic Stem Cell
HSS	DNase I Hypersensitivity Site
iCre	Improved Cre Recombinase
IMDM	Iscove's Modified Dulbecco's Medium
LCR	Locus Control Region
LIF	Leukaemia Inhibitory Factor
MAPK	Mitogen Activated Protein Kinase
MBP	Methylcytosine Binding Protein
MEP	Megakaryocyte Erythrocyte Progenitor
MOPS	3-(N-Morpholino) Propane-Sulphonic acid
MPP	Multi-Potent Progenitor
mRNA	messenger Ribonucleic acid
NLS	Nuclear Localisation Signal
OD	Optical Density
ORF	Open Reading Frame
PAC	P1-derived Artificial Chromosome
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction

PE	Phycoerythrin
PECY7	PE and a cyanine dye (Cy7)
PEV	Position Effect Variegation
pfu	plaque forming units
PI	Propidium Iodide
Pipes	Piperazine-N,N'-bis (2-ethanosulphonic acid)
PMSF	Phenylmethylsulfonyl Fluoride
Rb	Retinoblastoma
RNA	Ribonucleic Acid
rpm	rounds per minute
RPMI	Roswell Park Memorial Institute 1640 medium
RT	Room Temperature
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SP	CD4+CD8- or CD4-CD8+ Single Positive
SSC	Saline-Sodium Citrate
TAE	Tris Acetate (EDTA) Electrophoresis Buffer
TBE	Tris Borate (EDTA) Electrophoresis Buffer
TCR	T Cell Receptor
TE	Tris EDTA Buffer
TEMED	N,N,N',N',-Tetramethylethylenediamine
Tris	Tris (hydroxymethyl) Aminomethane
UTR	Untranslated Region
UV	Ultra Violet
V	Volts

v/v	volume per volume
w/v	weight per volume
ΔHMG	Truncated version of HBP1 without the HMG box

Publications

Some of the work described in this thesis resulted in two publications:

de Boer, J., Williams, A., Skavdis, G., Harker, N., Coles, M., Tolaini, M., Norton, T., Williams, K., Roderick, K., Potocnik, A. J., and Kioussis, D. (2003). Transgenic mice with hematopoietic and lymphoid specific expression of Cre. Eur J Immunol 33, 314-325.

de Chiara, C., Giannini, C., Adinolfi, S., de Boer, J., Guida, S., Ramos, A., Jodice, C., Kioussis, D., and Pastore, A. (2003). The AXH module: an independently folded domain common to ataxin-1 and HBP1. Febs Letters 551, 107-112.

Introduction

Introduction

Differentiation follows a common path in many different tissues and cell types. Expression of early differentiation markers is the first sign that precursor cells are undergoing a particular differentiation program. At the time of the cell exit, differentiating cells begin to induce expression of late tissue specific genes. Morphological changes can follow until fully differentiated phenotype is acquired. The differentiation process has been described in detail for many tissue types both *in vivo* and *in vitro.* These descriptions include cascades of gene induction and or repression and the morphological changes.

The orderly progression through all the differentiation steps is regulated at the molecular level. The early differentiation genes expressed in committed precursors are in most cases tissue specific transcription factors required for activation for later genes that are essential for full differentiation or repression of genes that were essential for the precursor but incompatible with full differentiation. The regulation of the factors is a very complex process requiring the coordination of different cellular pathways and molecular events. The interaction between trans-acting transcription factors and other protein co-factors with *cis* linked locus regulatory sequences represents the final level of gene activation or repression; that is, the assembly of a complete transcription or repressor complex. Many other events levels of control

precede these terminal events in the hierarchy of gene regulation, including cell cycle stage, integration of various extra-cellular signalling and signal transduction pathways and the physical state of a chromatin domain.

The retinoblastoma tumour suppressor protein pRB, and the closely related p107 and p130 proteins play important roles in many stages of differentiation process, including regulation of clonal expansion, cell cycle exit, induction of tissue specific gene expression and maintenance of the post mitotic state. pRB appears to regulate terminal exit from mitosis through at least two distinct mechanisms. The first follows directly from its role in regulating of the cell cycle and involves the repression of E2F transcriptional activity. The second is unique to the differentiation process and results from interaction with, for instance, HMG box containing protein 1 (HBP1), a factor that plays a role in cell cycle exit.

Mechanisms that inhibit cell cycle progression and establish growth arrest are fundamental to tumour suppression and to normal cell differentiation. This thesis will focus on the role of HBP1 in development, differentiation and cell cycle progression using a mouse model.

The introduction of this thesis will follow the same lines as described above. In the first part of this thesis, the haematopoietic development will be described as model for differentiation. The next part focuses on the molecular events underlying differentiation, such as the regulation of gene activation and repression. The cell cycle will be briefly described, followed by a description of the retinoblastoma tumour suppressor family and their function in the control of cell cycle progression and

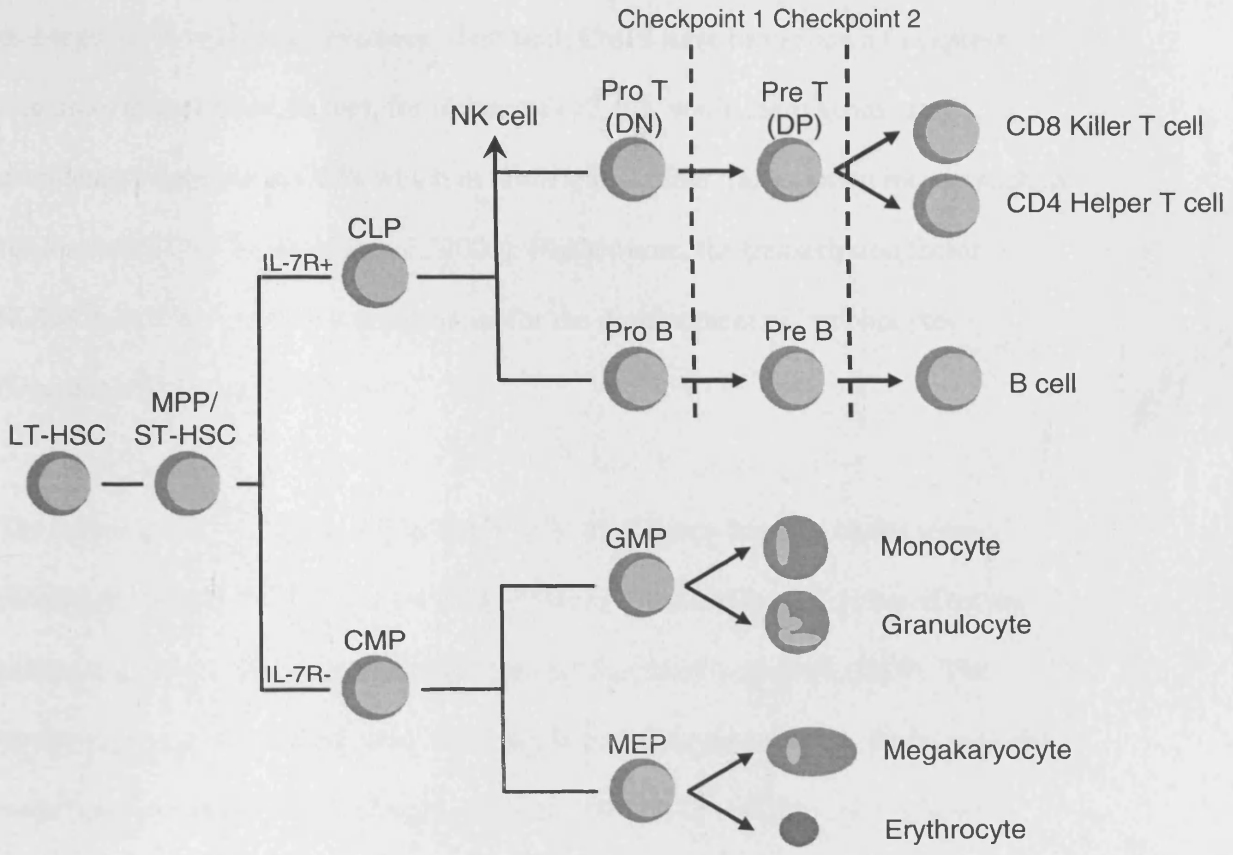
differentiation. Finally, the introduction will focus on HBP1 and associated partners and their roles and gene expression, cell cycle regulation and differentiation. A brief description on the gene targeting tools used for the work on HBP1 will also be given.

Development of the immune system

Haematopoietic stem cells (HSCs) are cells, which possess the properties of both self-renewal and multi-lineage potential for giving rise to all types of mature blood cells (see figure 1). Early HSC development displays a hierarchical arrangement starting with HSCs, which have extensive self-renewal capability. Next is the expansion stage, corresponding to proliferative multi-potent progenitor (MPP) or short term HSCs. MPP is also a stage of priming or preparation for differentiation. MPP then commits to either common lymphoid progenitor cells (CLP), which give rise to all the lymphoid lineages, or common myeloid progenitor cells (CMP), which produce all the myeloid and erythroid lineages. It is controversial whether this is the only pathway for lymphocyte development. Lineage determination of a differentiating stem cell may not be as rigidly defined as the above model predicts and committed progenitors may trans-differentiate when subjected to variety of micro-environmental regenerative cues (Blau et al., 2001).

Figure 1: Model of major haematopoietic maturation pathways from HSC.

Long Term Haematopoietic Stem Cell (LT-HSC) gives rise to Multi-Potent Progenitor (MPP) or short term HSCs. The MPP give rise to Common Lymphoid Progenitor (CLP), which can form all cells of the lymphoid lineage, and Common Myeloid Progenitor (CMP), which can differentiate in either Granulocyte Monocyte Progenitor (GMP) or Megakaryocyte Erythrocyte Progenitor (MEP), which form the cells of the granulocyte/macrophage or megakaryocyte/erythroid lineages, respectively. From the CMP both B and T cells can be derived. The figure shows the timing in both lineages of checkpoints at which the population undergoes rigorous selection that eliminates many or most of the developing cells (Checkpoints 1 & 2) based on properties of their antigen receptor gene products. In the T-cell pathway, Double Negative (DN) Pro-T cells turn on both CD4 and CD8 to become a double positive (DP) Pre-T cell but ultimately down regulate one or the other if they succeed in passing checkpoint 2. Expression of CD4 and CD8 is mutually exclusive and stable thereafter.


Checkpoint 1 Checkpoint 2
NK cell
Pro T (DN)
Pre T (DP)
CD8 Killer T cell
CD4 Helper T cell
CLP
IL-7R+
Pro B
Pre B
B cell
MPP/ ST-HSC
LT-HSC
GMP
Monocyte
Granulocyte
CMP
IL-7R-
MEP
Megakaryocyte
Erythrocyte

The main difference between the CLP and the CMP is the expression of interleukin-7-receptor (IL-7R), CLPs express the IL7R on the cell surface while this receptor is absent on CMPs (Sudo et al., 1993). The exact mechanism of what makes a HSC commit to either CLPs or CMPs is still unknown. However, key transcription factors that regulate this process have been identified. CMPs have been shown to express important transcription factors, for instance GATA-1, while these genes are completely repressed in CLPs which in turn express other transcription factors such as Aiolos and GATA-3 (Akashi et al., 2000). Furthermore, the transcription factor IKAROS has been shown to be essential for the development of lymphocytes (Georgopoulos et al., 1992).

The commitment of CLPs to either the B or T cell lineage depends on the trans-membrane receptor Notch-1. A model proposed by Bussinger et al. is based on the assumption that Notch-1 is expressed on all CLPs (Busslinger et al., 2000). The thymic environment contains Notch-1 ligands and upon migration to the thymus, the model proposes that CLPs will be subjected to Notch-1 activation, inducing the progenitors to develop along the T-cell lineage. However, there are no Notch-1 ligands in the bone marrow and hence the CLPs cannot be activated through the Notch-1 receptor. The absence of Notch-1 therefore allows the CLPs to differentiate to the B cell lineage and up regulate B cell specific transcription factors E2A and EBF which in turn activate Pax5 which represses certain genes and thereby commits the cells to the B cell lineage (Busslinger et al., 2000).

B and T lymphocytes undergo developmental pathways that resemble each other in many respects. Both B and T cells express clonally unique cell surface antigen receptors. These receptors are encoded by genes that are assembled by a developmentally programmed mechanism of DNA rearrangements. Antigen receptor (immonuglobin [Ig] for B cells or T cell receptor [TCR] for T cells) gene rearrangement process is controlled by expression of the recombinase genes RAG1 and RAG2 on both T and B cells. During lymphocyte development the developing B or T cell is subjected to several checkpoints. At these checkpoints most cells are eliminated and the few successful will proliferate strongly. The rearrangement of the gene encoding for the first chain of the antigen receptor forms the first checkpoint. All the cells in which the rearrangement was successful will proliferate and pass to the next checkpoint. At this checkpoint the functionality of the antigen receptor is determined. Only cells with an antigen receptor with an appropriate recognition specificity to be useful for the immune system will be allowed to become functional T or B cells. The developmental stages of the developing lymphocytes can be measured by the expression of certain cell surface markers.

Gene regulation

From the above it is clear that during lymphocyte differentiation specific transitions in gene expression are necessary for full differentiation. Gene expression in eukaryotes is regulated primarily at the level of transcription (Latchman, 1997). The level of transcription for many genes important for cell proliferation and differentiation is regulated in turn at the chromatin level. In this chapter a description of chromatin will

be given and how packaging by chromatin influences gene transcription. The key proteins and sequence motifs that facilitate in these transitions between chromatin states will also be described. In the last part of this chapter studies on the regulation of hCD2 locus will be described as an example of a gene expressed during haematopoiesis.

Chromatin

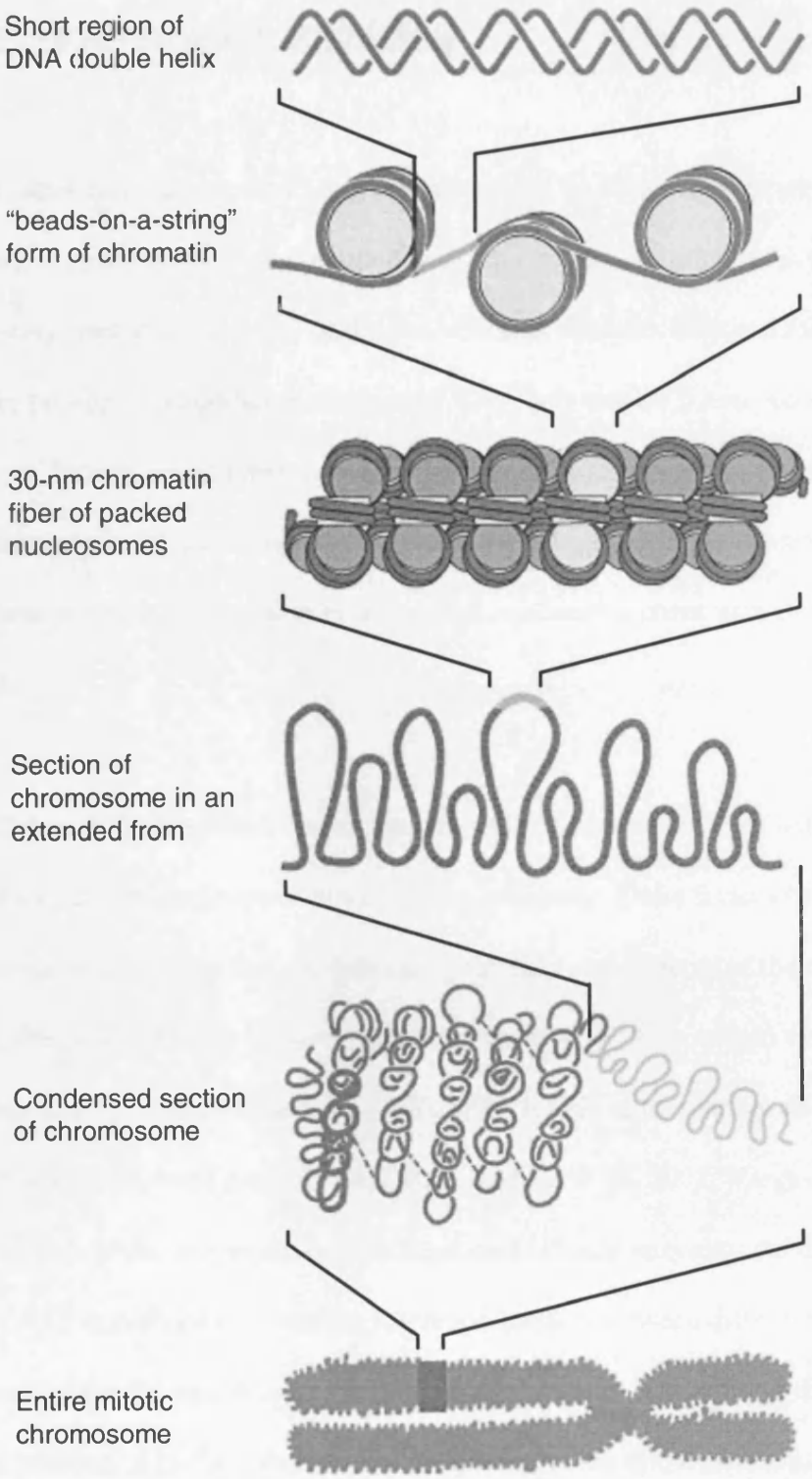
In eukaryotes, genomic DNA is packaged by histones and non-histone proteins into chromatin. The histone proteins fall into five major classes H1, H2A, H2B, H3 and H4. Each histone class includes some gene variants or subtypes and is likely to provide tissue specific and developmental stage specific variation of the chromatin structure (Doenecke et al., 1997). Histones are an essential part of the fundamental repeating unit of chromatin: the nucleosome. The nucleosome core contains about 147 bp of DNA and two copies each of the histones H2A, H2B, H3 and H4. The additional linker DNA that connects the nucleosome cores is of variable length. The nucleosome core, linker DNA and histone H1 make up the complete nucleosome that can be visualized with electron microscopy as the classic "beads on a string" appearance (figure 2). However, this is only the first step in the compaction process that the DNA undergoes to fit in the nucleus. The next step in higher order chromatin structure is the so called 30nm fibre which is made when the "beads on a string" winds into a regular coil (figure 2). Linker histones, non-histone proteins and potential interactions of core histone N-terminal tails with adjacent nucleosomes contribute to

this compaction of nucleosomal DNA. In the interphase nucleus two types of higher order structures can be distinguished; the highly condensed regions referred to as heterochromatin and the less condensed/decondensed regions referred to as euchromatin. Finally, the highest level of packaging occurs in the M phase of the cell cycle when the fibre is organized in loops and domains that compact the DNA by more than 10,000 fold. These can be visualised as chromosomes (figure 2).

It is thought that dynamic changes in chromatin structure facilitate or prevent the access of transcription factors to nucleosomal DNA, which in turn affect the expression of genes. Inaccessibility of a region to the transcriptional machinery, will lead to silencing of the genes that lie within this region. For transcription to proceed the structure of chromatin has to be unfolded and inaccessibility be overcome. At least two mechanisms can be used to alter or remodel chromatin structure. One mechanism involves multi-subunit protein complexes that use ATP as a source of energy, which alter chromatin structure by changing the location or conformation of the nucleosome (Becker and Horz, 2002). These structural changes are accomplished without covalent modification, and can be involved in either activation or repression (Varga-Weisz, 2001). A second mechanism involves covalent modifications either of DNA (Bird and Wolffe, 1999) or of histone N-terminal tails that protrude from the core nucleosome (Zheng and Hayes, 2003). In the following paragraphs, both kinds of modification will be discussed. It has to be noted however that both types of modification appear to work in concert in order to establish or maintain chromatin in either an active or silent state.

Figure 2: Chromatin structure

The eukaryotic chromosome exhibits many levels of chromatin compaction, starting with the repeating nucleosomal unit, which wraps DNA on the surface of the histone octamer, and ending with the highly condensed and folded structure that characterizes the mitotic chromosome. (Adapted from Alberts et al. (1998), Essential Cell Biology)


Short region of DNA double helix
"beads-on-a-string" form of chromatin
30-nm chromatin fiber of packed nucleosomes
Section of chromosome in an extended from
Condensed section of chromosome
Entire mitotic chromosome

ATP dependent chromatin remodeling

ATP dependent chromatin remodelling complexes can be involved in nearly every step of gene regulation. They can regulate key steps in transcriptional activation before, during, and after assembly of the pre-initiation complex (Fry and Peterson, 2002). The pre-initiation complex, containing RNA polymerase II and other general transcription factors, is assembled at the target gene promoter prior to transcription. Besides transcriptional activation ATP dependent chromatin complexes can also be involved transcriptional repression by setting up inaccessible chromatin structures at promoters.

Many different ATP-dependent chromatin remodelling complexes have been identified which alter nucleosome structure or positioning. These factors have been separated into two families that are defined by the different features of their ATPase subunits. The 'ISWI' family of complexes all contain an ATPase related to *Drosophila* ISWI protein, whereas the "SWI/SNF" family of complexes all contain a protein related to the yeast protein Swi2/Snf2 (Narlikar et al., 2002; Varga-Weisz, 2001). These families of complexes both function as classic enzymes and use the energy of ATP hydrolysis to lower the activation barriers between different nucleosomal states (Varga-Weisz, 2001 2). These chromatin states might differ in either the position of nucleosomes and/or the conformation of the nucleosome. Whether the altered nucleosome structure is transcriptionally active or inactive depends on the transcription factors that target these ATP-dependent chromatin

remodelling complexes to the DNA. If transcriptional repressors are involved, the remodelling complexes will change the chromatin conformation into a repressed state. Conversely, involvement of transcriptional activators will change the chromatin conformation to an active state.

Covalent chromatin modifications

Covalent modifications can take place on the DNA itself or on the histone tails. Both can have profound effects on gene expression and DNA modifications will be discussed first followed by the modifications on the histone tails.

DNA methylation

DNA can be methylated at position 5 of cytosine within CpG dinucleotides, a modification that is associated with transcriptional repression; this methylation influences chromatin structure and as a consequence gene activity (Bird and Wolffe, 1999). This modification is carried out by methyltransferases (DNMTs). Sequences that include methylcytosines are specifically recognized by Methylcytosine Binding Proteins (MBPs), such as MeCP2 and MBPs 1/2/3. These proteins have been shown to be transcriptional repressors (Bird and Wolffe, 1999). MECP2, the best studied MBP, will be described as an example of how MBPs can function. MECP2 contains two domains: the methyl-CpG binding domain, which recognizes a methylated-CpG dinucleotide through contacts in the major groove of the double helix and a

transcriptional repression domain, which interacts with the SIN3 complex. The SIN3 complex contains two histone de-acetylases and two histone-binding proteins (Ahringer, 2000). This led to a model that predicts that MeCP2 is recruited to methylated regions of the genome where its interactions with a protein complex containing Sin3 and histone de-acetylases lead to the establishment and maintenance of repressive chromatin architecture (Bird and Wolffe, 1999).

Histone tail modification

Histone tails can be modified in several ways, including acetylation, methylation ubiquitination and phosphorylation. In the following paragraphs, these modifications will be described briefly.

Acetylation

The acetylation status of lysine residues in the core histone tails is under control of nuclear Histone Acetyl-Transferases (HATs) and of Histone De-acetylases (HDACs). Several mammalian HATs and HDACs have been characterised to date. These enzymes frequently targeted the highly conserved histone H3 lysines at amino-terminal amino-acid positions 9, 14, 18 and 23, and H4 lysines 5, 8, 12 and 16 for modification (Roth et al., 2001). A general correlation between histone acetylation and gene activity in vivo is now well established, although there are some exceptions (Eberharter and Becker, 2002). Lysine acetylation results in an altered charge density

of the tail domains. Neutralization of the positive charge of lysine residues by acetylation is thought to weaken the interactions between the histone tails and the negatively charged DNA backbone leading to a "looser" or more "open" chromatin configuration for the entry for transcription factors (Morales and Richard-Foy, 2000). It also provides a binding site for bromo domains that can recognize acetyl-lysine (Zeng and Zhou, 2002). Bromo domains are found in several eukaryotic transcription factors, including most nuclear histone HATs (Jeanmougin et al., 1997) and all known SWI/SNF-like complexes (Horn and Peterson, 2001). Thus, lysine acetylation in the tails of histones has been associated with actively transcribed or poised chromatin; the binding of bromo domains to the acetylated lysines allows the anchoring of HATs and other co-activators onto active chromatin.

Methylation

There are two types of histone methylation, targeting either arginine or lysine residues within the tails of histone H3 or H4. Methylation of histones by arginine methyltransferases has been known for some time, but has been validated only recently as a functional event involved in gene activation. A breakthrough came when arginine methyltransferase CARM1 was shown to activate transcription via methylation of Histone H3 (Chen et al., 2002). Furthermore, estrogen-responsive gene expression was aberrant in $Carm1^{-/-}$ fibroblasts and embryos, thus emphasizing the role of arginine methylation as a transcription activation tag (Yadav et al., 2003).

Methylation of histones on lysine residues has also been implicated in gene regulation. The lysines at the fourth (K4) and ninth (K9) positions on the tail of the histone H3 can regulate gene expression when a methyl group is added to these two positions. Generally, the methylation of H3K4 correlates with gene expression but exceptions to this model have been described (Sun and Allis, 2002). Stretches of chromatin in which histone H3 is methylated on position nine (H3K9) correlates with transcriptional repression by establishing regions of transcriptionally inert heterochromatin (Lachner and Jenuwein, 2002). This idea is supported by several recent discoveries. Methyltransferases SUV39 (mammalian homologue of Drosophila Su(var)3-9), that specifically methylates histone H3K9, associates with the heterochromatin protein HP1. HP1 is a conserved chromosomal protein that participates in restrictive chromatin packaging and gene silencing (Kouzarides, 2002; Zhang and Reinberg, 2001). Based on these findings, mechanisms for heterochromatin spreading have been proposed, involving recruitment of SUV39 by HP1 and propagation of the methylated mark along the chromosome via methylation of histone H3K9.

Phosphorylation

Histone phosphorylation involving serine residue position 10 of histone H3 (H3S10) has also emerged as an important modification, both in transcriptional activation and in chromosome condensation during mitosis (Cheung et al., 2000). Further studies suggest that phosphorylation of H3S10 facilitates acetylation of H3K9 and/or H3K14

(Cheung et al., 2000) whereas H3S10 phosphorylation is suppressed if methylation of H3K9 occurs first (Rea et al., 2000).

Ubiquitination

Recent developments indicate that histone ubiquitination is another important modification. In yeast Saccharomyces cerevisiae, for example, addition of ubiquitin on H2BK123 is critical to mitotic and meiotic cell division (Shekhar et al., 2002). Moreover, recent studies in yeast show that ubiquitination of histone H2B controls the outcome of methylation of H3K4. The pathway described here results in transcriptional silencing (Sun and Allis, 2002), even though H3K4 methylation is generally regarded as an activation mark for chromatin remodelling.

Cis acting sequences

The above described chromosomal modifications do not happen randomly in the genome, but are targeted to certain regions which are determined by their nucleotide sequences. In the next paragraph, some of these cis acting sequences/regions will be described.

Eukaryotic genes are arranged into coding and non-coding DNA sequences, exons and introns respectively. In addition to the protein coding DNA, other important extra-genic DNA sequences are present which are involved in gene regulation. The

three main types of regulatory sequences that have been defined are promoters, enhancers and silencers, but additional elements exist, such as boundary elements and Locus Control Regions (LCRs). All these elements mediate their function by binding protein factors that recognise specific DNA sequences. The presence of a factor combined with its specificity for certain regulatory elements ensures the differential expression of genes in different cell types and developmental stages.

The gene promoter

The core promoter is defined as the minimal stretch of contiguous DNA sequence that is sufficient to direct accurate initiation of transcription by the RNA polymerase II machinery (Butler and Kadonaga, 2002). Core promoters comprise DNA sequence motifs within 40 nucleotides upstream and downstream of the RNA start site that, in the appropriate combinations, are sufficient to direct transcription initiation by the basal RNA polymerase II transcriptional machinery (Butler and Kadonaga, 2002). RNA polymerase II is a multi-subunit enzyme that catalyzes the synthesis of mRNA from the DNA template (Butler and Kadonaga, 2002). Accurate and efficient transcription from the core promoter requires the polymerase along with auxiliary factors that include (TF)IIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Butler and Kadonaga, 2002). In the stepwise RNA polymerase II complex assembly TFIID and TFIIB are the first two factors that interact with the core promoter (Butler and Kadonaga, 2002).

Enhancers and Silencers

Transcriptional enhancers are DNA sequences that control expression of nearby genes in a positive manner (Blackwood and Kadonaga, 1998). The general mode of action of these sequences is that they associate with specific activator proteins; this association of enhancer and activator facilitates transcription initiation at the target gene's promoter. Enhancers facilitate in the initiation of transcription through direct contact between the enhancer and promoter. Proof is this came when it was found that the DNase I hypersensitive sites of the β-globin locus (these contain the enhancers of the β-globin locus) and an active globin promoter are held in close physical proximity when the gene is transcribed (Carter et al., 2002). As yet, it is unclear how these loops are formed. It is possible that enhancer and promoter create a loop without interacting with the intervening DNA. It is also possible that components bound to the enhancer track along the chromatin towards a promoter that eventually results in a loop formation.

Enhancers might be located upstream, downstream or even within the transcription unit that they control and their action appears to be orientation independent (Blackwood and Kadonaga, 1998). Under the influence of one or more enhancers, a gene can be abundantly expressed in certain tissues of the organism (i.e., cells in which the activator protein is found) and weakly or not at all expressed in other tissues (Blackwood and Kadonaga, 1998).

Silencers act similarly to enhancers in that they affect gene promoters from a distance in a position and orientation independent manner. However, the function of the silencer, as its name implies, is to down-regulate gene expression.

Boundary elements

Boundary elements or insulators are thought to separate chromatin into distinct units or domains with differential gene regulation (Bell and Felsenfeld, 1999). Boundary elements, when inserted between an enhancer and a promoter, can inhibit the enhancer from activating a promoter. The enhancer-blocking activity of boundary elements is clearly distinct from silencing, because it is strictly dependent on its position and occurs without repressing either the enhancer or the promoter (Bell and Felsenfeld, 1999). Furthermore, boundary elements have been described to insulate a transgene from position effects (Bell and Felsenfeld, 1999).

Factors affecting gene expression levels: Position effect variegation

The effects of the chromatin environment on gene expression is best seen when genes are translocated from their own normal location to a different chromosomal location. This is most often seen when transgenic animals are made (Kioussis and Festenstein, 1997). Several factors conspire to make the expression patterns of transgenes highly unpredictable when transgenic animals are made. It is thought that these unpredictable expression patterns are caused by variations in the site of integration, the so called

position effects, and that these variations are, at least in part, due to differences in the nature of the surrounding chromatin (Kioussis and Festenstein, 1997). These effects are exemplified by a phenomenon called Position Effect Variegation (PEV).

PEV can be defined as a position-dependent inactivation of gene expression in a fraction of a cell population of the same lineage (Henikoff, 1990). One of the best studied instances of PEV is a chromosomal rearrangement in Drosophila in which an allele of the white gene is transferred to a site close to the centromere (Wallrath and Elgin, 1995). After this translocation its previously uniform expression becomes "variegated," producing patches of pigmented and unpigmented cells in the eye (Wallrath and Elgin, 1995). It is thought that a peri-centromeric location renders the gene susceptible to the variable spreading of heterochromatic condensation (Wallrath and Elgin, 1995). PEV has also been demonstrated in yeast (Allshire et al., 1994) and in mammals (Festenstein et al., 1996) for genes located within or near centromeres or telomeres.

Variegated expression has been observed in transgenes as well as in endogenous genes (Kleinjan and van Heyningen, 1998). As the expression level of a transgene is highly dependent on its integration site, which is not predetermined when transgenic animals are made, the forces leading to PEV can cause large and unpredictable variations in expression levels. In the case of mice, it has been reported that transgene integration into peri-centromeric regions is the most frequent inactivating process (Festenstein et al., 1996).

Locus Control Region

In some remarkable cases transgenes are expressed in an integration-site independent manner (Grosveld et al., 1987). Attempts to identify the responsible elements for this protective effect have let to the discovery of locus control regions (LCR) (Grosveld et al., 1987). LCRs are operationally defined by their ability to enhance the expression of linked genes to physiological levels in a tissue-specific and copy-number dependent manner at ectopic chromatin sites (Grosveld et al., 1987). The components of an LCR commonly co-localize with sites of DNase I hypersensitivity (HS) in the chromatin of expressing cells (Li et al., 2002). The core determinants at individual HSs are composed of arrays of multiple ubiquitous and lineage-specific transcription factor-binding sites.

The existence of LCRs was first identified in the human β-globin locus (Grosveld et al., 1987). Evidence for the presence of an LCR came from transgenic mouse studies. Linkage of this region to a β-globin gene resulted in expression of the gene at a level comparable to the endogenous mouse β-globin genes in a position-independent, copy number-dependent manner (Grosveld et al., 1987). LCRs have been described in a broad spectrum of mammalian gene systems, suggesting that they play an important role in the control of eukaryotic gene expression (Li et al., 2002). In the following chapter, the studies on the hCD2 LCR will be described as a model of LCR function.

Human CD2 Locus

The human CD2 molecule is expressed on virtually all T cells and thymocytes as well as NK cells. Human CD2 binds to the ubiquitous CD58 (LFA-3) cell surface glycoprotein, and through this receptor interaction, promotes the initial stages of T cell contact with antigen presenting cells (Springer, 1990) providing a mechanism for sustained TCR engagement and signalling and thereby optimizing T cell activation (Bierer et al., 1988). Indeed, CD2 is essential for the immune response against weak agonistic ligands (Sasada and Reinherz, 2001). Thus, CD2 sets a quantitative threshold for T cell activation. It has to be noted that the CD2 gene is diverged during evolution to recognize CD58 in humans and CD48 in rodents. The interaction with human CD2 and murine CD48 have not been demonstrated (van der Merwe et al., 1994) indicating that human CD2 is inactive in a mouse model.

To study the mechanisms underlying CD2 expression transgenic animals were made with a 29.4 kb *Kpn I* fragment containing the human CD2 locus (figure 3A). This fragment contains 5.3 kb of DNA upstream from the 5' end of the gene and 9.3 kb downstream from the second poly(A) addition site. Analysis of transgenic mice carrying the human CD2 29.4 kb *Kpn I* fragment showed that the levels of expression of the transgene was dependent on the copy number of the genes integrated into the mouse genome (Lang et al., 1988). These experiments showed that the hCD2 locus contains an element, which is capable of dominantly inducing changes in chromatin structure that are required for the expression of the genes, or alternatively, protect the gene from hetero-chromatinisation. As mentioned before locus control regions are functionally defined in transgenic mouse experiments as elements that give rise to full

levels of expression of a linked transgene independent of the site of integration in the host genome (Grosveld et al., 1987). From this, it was concluded that the hCD2 locus contains a locus control region.

To identify further the sequences responsible for the copy number dependent expression, deletions were made in the 29.4 kb *Kpn I* fragment. Removal of sequences downstream of the second poly-adenylation site of the CD2, which left the promoter and gene intact, resulted in gene silencing. It was concluded that this 3' region was necessary for the appropriate regulation of the expression of the CD2 gene (Greaves et al., 1989).

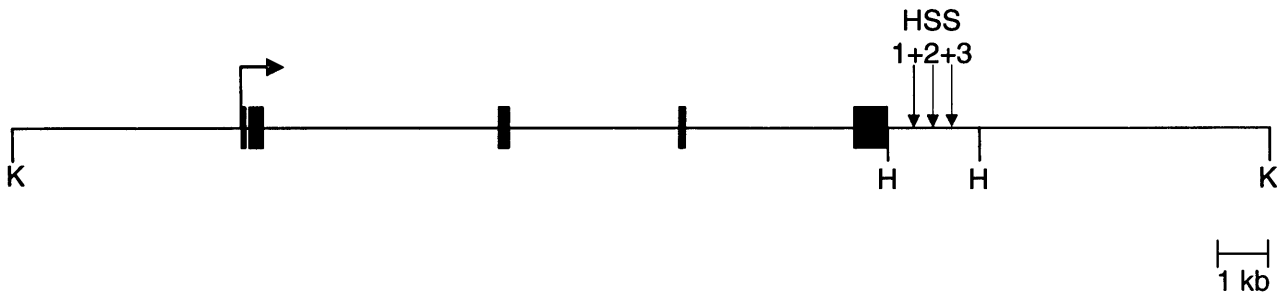
The 3'flanking sequences were further characterized by a range of deletions. Fragments covering 0.2, 0.5, 1.1, 1.3, 1.5 and 2.1 kb of 3' flanking sequences were made and linked to a hCD2 mini gene. This showed that the minimum sequences needed for copy number dependent and position independent expression are 2kb of 3' flanking sequence (figure 3B) (Festenstein et al., 1996; Lang et al., 1991). It is reasonable to hypothesize that a number of regulatory proteins would bind to this 2kb fragment and that interactions between these regulatory proteins, leading to the formation of multi-molecular complexes, insures copy number dependent and position independent expression of the hCD2 locus. Thus, in order to characterize how an LCR functions it is of interest to identify the necessary regulatory proteins that would bind to this 2kb fragment.

Figure 3: Human CD2 locus

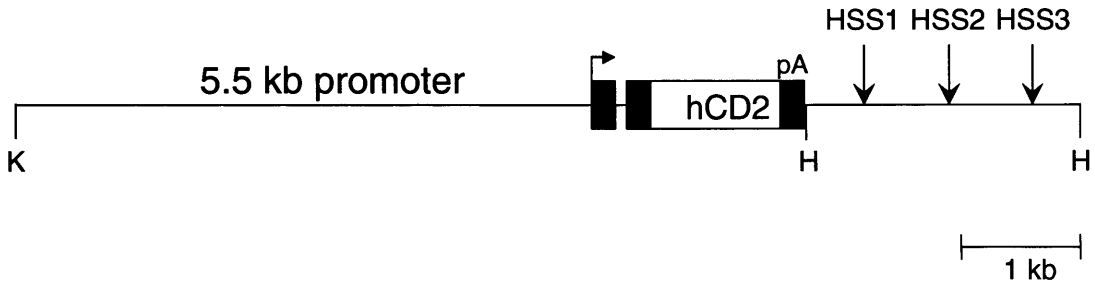
A) The human CD2 locus is contained in a 29.4 kb *Kpn I* (K) fragment. This fragment contains 5.3 kb of DNA upstream from the 5' end of the gene and 9.3 kb downstream from the second poly(A) addition site. Three identified HSS 3' of the gene are indicated. The exons are indicated as black bars.

B) The hCD2 mini gene contains 5.5kb of the promoter region, the human CD2 gene with all but the first intron deleted and the minimal LCR, a 2.1 kb *HindIII* (H) fragment 3' of the poly(A) site). The LCR region contains a classical enhancer (HSS1) and a region without enhancer capacity but necessary for LCR function (HSS3).

A) hCD2 genomic locus



B) hCD2 mini gene



It is thought that multi-molecular complexes formed at the enhancer or promoter sequences result in the remodelling of chromatin at these points and in the formation of nuclease hypersensitive sites (HSS). Three clusters of such sites were identified within 2 kb of 3'flanking region of the hCD2 gene. Clusters HSS1 and HSS3 are shown in figure 3B (Festenstein et al., 1996; Greaves et al., 1989; Lang et al., 1991).

Within the 2kb of 3'flanking region the upstream HSS cluster of strong hypersensitive sites (HSS1) coincides with the region that functions as a classical enhancer, whereas the weak HSS2 and HSS3 have no activity in transient transfection assays (Lake et al., 1990). The role of HSS3 was further investigated by generating transgenic mice carrying constructs lacking HSS3 (Festenstein et al., 1996). Removal of HSS3 resulted in loss of position independence and led to mosaic expression rather than a general lower expression (Festenstein et al., 1996). This mosaic expression is thought to be the result of a stochastic process and once the decision is made to express or to silence the human CD2 transgene, it is maintained through subsequent cell divisions (Festenstein et al., 1996). The phenomenon is a typical example of position effect variegation (PEV) and is characterised by the 'stochastic' silencing of the hCD2 transgene in a proportion of cells within the T cell lineage that would normally be expected to express it (Festenstein et al., 1996; Henikoff, 1990).

Non expressing cells show a closed chromatin configuration of the transgenic CD2 locus as illustrated by a lack of DNase I hypersensitivity compared to cells that still express the human CD2 gene (Festenstein et al., 1996). Subsequently, it was shown that PEV of the disabled LCR constructs is sensitive to chromatin modifiers, like HP1-β, in a dose dependent manner. Thus, increased levels of HP1-β led to an

increase in the proportion of cells that silence the variegating transgene. On the other hand, the expression of a hCD2 transgene containing the full LCR was not affected by different levels of HP1-β (Festenstein et al., 1999).

In summary, the hCD2 LCR consists of at least two distinct elements: a transcriptional enhancer and a region without enhancer activity necessary for preventing heterochromatin induced position effect variegation. The latter region is thought to prevent PEV by ensuring an open chromatin configuration, so that other factors can bind to the rest of the LCR and the promoter of the human CD2.

In order to identify components that participate in the chromatin structures formed in this region, protein bindings sites within HSS3 were mapped using DNase I footprint analysis (Zhuma et al., 1999). Using one of the identified HSS3 footprints (FT1) as a bait in a yeast one hybrid assay led to the isolation of HMG Box containing Protein 1 (HBP1) (Zhuma et al., 1999). It was shown subsequently that HBP1 binds FT1 in a sequence specific manner (Zhuma et al., 1999). Deletion of the binding site of HBP1 (FT1) from the hCD2 transgene results in its variegated expression when the transgene is integrated close to heterochromatic areas (Zhuma et al., 1999). Thus, it was concluded that HBP1 contributes to the hCD2 LCR function in preventing PEV. The contribution of HBP1 to the establishment of an active chromatin structure was confirmed by over-expression of a transgenic HBP1 in developing thymocytes. Thus, over-expressing HBP1 mice were crossed to a hCD2 variegating mouse line that has 1.5 kb of the LCR. The 1.5 kb of the LCR still has the binding site for HBP1 (FT1), but is missing the rest of HSS3 and, thus, exhibits PEV. Over-expression of HBP1 in these hCD2 transgenic mice increases the number of cells that express hCD2

compared to the transgenic mice with only endogenous levels of HBP1. Furthermore the variegation seen in transgenic mice lacking the FT1 sequences was unaffected by HBP1 over expression. These results suggest that HBP1 activates the CD2 locus (B. Sekkali submitted for publication). It is unclear at the moment how HBP1 activates the hCD2 locus. HBP1 may promote an active hCD2 locus by binding to and bending the DNA, thus allowing DNA-protein and/or protein-protein interactions. It is also possible that the phenotype seen is through the recruitment of chromatin remodelling complexes by HBP1. HBP1 can interact with the retinoblastoma that, in turn, can interact with a SWI/SNF complex (van der Merwe et al., 1994). Thus, the initial binding of HBP1 to hCD2 LCR may result in the recruitment of RB and subsequently chromatin remodelling complexes to the locus to activate transcription.

The cell cycle

Co-ordination and balance between proliferation and differentiation is required for normal development and for the maintenance of homeostasis in the adult.

All somatic cells proliferate using a fundamentally similar mechanism. This cell division cycle comprises a set of processes that allows the cell to produce a pair of genetically identical daughter cells. This involves a complicated control system that has been conserved across evolution. Many of the proteins, which regulate the cell cycle in a human cell, are able to function when transferred to yeast. A eukaryotic cell cycle is traditionally divided into several distinct phases (figure 4A). During G1 cells

make the necessary preparations for DNA synthesis, which occurs during S phase. Likewise, during G2 cells prepare for mitosis (M phase), during which replicated genetic material is segregated into daughter cells. Before the cells enter the mitotic cell cycle, they are in a state of quiescence. This state is often referred to as G0. Quiescent cells remain competent to respond to growth factors by progressing through G1 (Cook et al., 2000).

To maintain homeostasis, a tight control of the cell cycle process is required. Failure of cell cycle regulation can have a wide variety of consequences, such as incorrect number of cells produced during development or deregulated growth of terminally differentiated cells. Regulation of cell division is dependent upon various extra cellular signals, which include growth and differentiation factors, survival factors and cellular stresses (Cook et al., 2000). These extra cellular cues activate signal transduction pathways which induce programmes of gene expression and ultimately determine whether the cell enters the cell cycle and divides, leaves the cell cycle and differentiates, or undergoes apoptotic cell death (Cook et al., 2000).

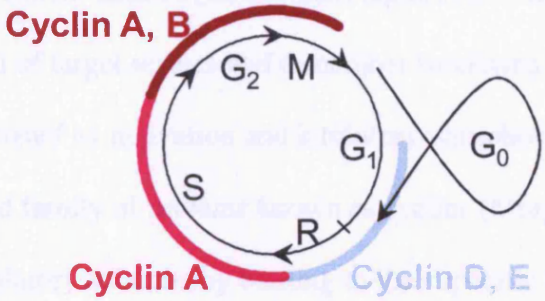
Upon mitogenic stimulation, cells enter the G1 phase. Progression through G1 is the major growth-factor-responsive cell cycle transition; once the cell reaches a point late in G1, termed the 'restriction point', it becomes committed to proceed through S phase as far as G2/M.

Figure 4: The eukaryotic cell cycle

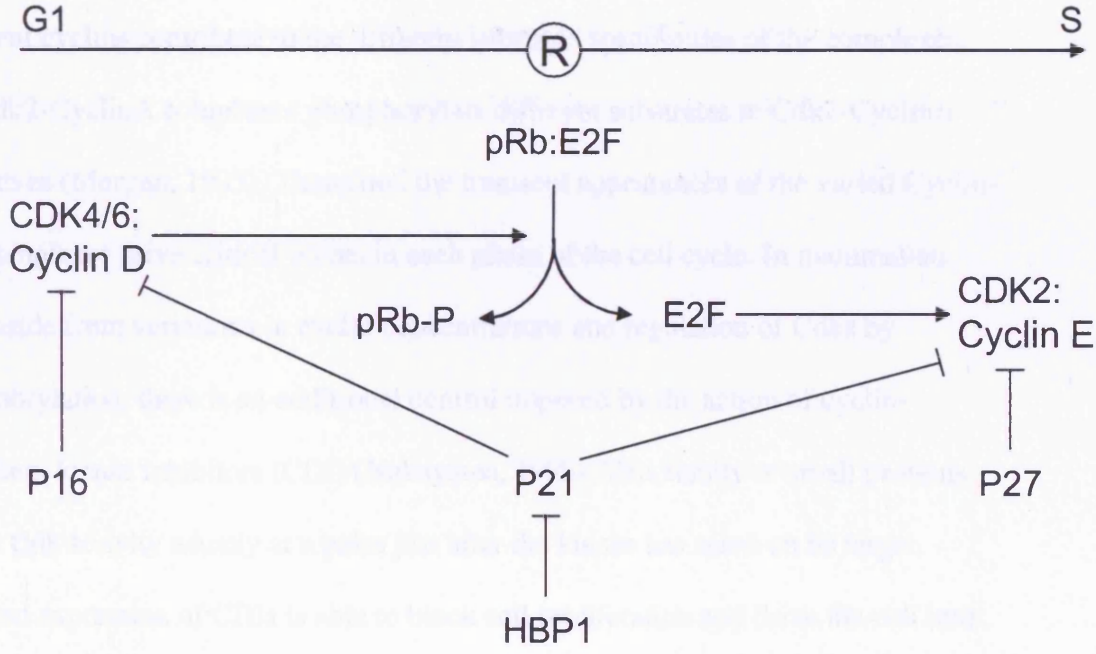
A) The successive phases of a standard eukaryotic cell cycle. See text for details.

B) G1/S transition. This model represents key steps in the initiation of the G1/S cell cycle transition by cyclin D1 CDK4/6 complex and the subsequent activation of cyclin E by Retinoblastoma. P16, P21 and P27 are negative regulators of the G1/S transition. The restriction point is indicated with R.

A: Overview of cell cycle



B: Overview of G1 to S transition



An intricate control system oversees the timely and ordered succession of the various phases. After DNA replication, there is another major checkpoint to ensure the genome has been properly replicated before entering the M phase. The control system that oversees the cell cycle is based primarily on two key protein families. The first is the family of cyclin-dependent kinases (Cdk) which regulate downstream proteins through phosphorylation of target serines and threonines (reviewed in MacLachlan *et al*, 1995). Cdks are regulated by activation and inhibitory phosphorylation, and by interaction with a second family of proteins known as cyclins (Morgan, 1995). Cyclins function as regulatory subunits by binding to their specific Cdk molecule and contribute to the activation of the kinase. Once activated, the complexes are able to phosphorylate specific nuclear targets. Regulation is achieved through synthesis and degradation of the cyclin molecules, as the cyclical assembly, activation and dismantling of these Cyclin-Cdk complexes are pivotal events during the cell cycle. Different cyclins contribute to the different substrate specificities of the complexes, e.g. Cdk2-CyclinA complexes phosphorylate different substrates to Cdk2-CyclinE complexes (Morgan, 1995). Therefore, the transient appearances of the varied Cyclin-Cdk complexes drive critical points in each phase of the cell cycle. In mammalian cells, aside from variations in cyclin concentrations and regulation of Cdks by phosphorylation, there is an additional control imposed by the action of cyclin-dependent kinase inhibitors (CDI) (Nakayama, 1998). This family of small proteins inhibit Cdk activity usually at a point just after the kinase has acted on its target. Elevated expression of CDIs is able to block cell proliferation and drive the cell into quiescence. In addition, CDIs are also involved in promoting differentiation, maintenance of the post-mitotic state in terminally differentiated cells and response of the cell cycle machinery to genotoxic stress. The CDIs are divided into two classes,

based on their sequence similarity and specificity of action. One is the Cip-Kip family, which are capable of all of the kinases relating to the G1 phase (Cdk2, Cdk3, Cdk4 and Cdk6). Members of this family include p21, p27 and p57. The other class comprises the INK4 family and the main targets of these CDIs are Cdk4 and Cdk6. This family is composed of p15, p16, p18 and p19. In the following chapter, the mechanisms of the regulation of Cyclin/Cdk expression and activation during the G1 to S phase will be described in more detail.

G1-S transition

As shown before, cell cycle checkpoints are controlled by cyclin dependent kinases (CDKs). The CDKs act at key steps in the mammalian cell cycle to effect an orderly transition between stages. CDKs are activated by binding to cyclins (Cook et al., 2000).

Following mitogen stimulation of quiescent cells, genes encoding D-type cyclins are expressed at the beginning of the G1 phase. Cyclin D then binds CDK4 or CDK6 (figure 4B). The binding of cyclin D with CDK4/6 creates an active complex. This active cyclin D:CDK complex initiates cell cycle progression towards S phase. Cyclin D:CDK4/6 complex phosphorylates the retinoblastoma tumour suppressor protein (pRB) which results in its inactivation leading to progression from G1 to the S phase of the cell cycle. The growth-suppressive function of pRB lies in its ability to bind and sequester the sequence-specific transcription factor E2F (Harbour and Dean, 2000). Inactivation of pRB after phosphorylation by cyclin D:CDK releases E2F

(Coqueret, 2002). Subsequently, E2F promotes the expression of a variety of genes required for transit through late G1 and entry into S phase, including cyclin E, which is needed for the next step in the cell cycle (Coqueret, 2002) (figure 3).

The active cyclin-CDK complexes are inhibited by CDK inhibitor proteins (CDKIs) (Harper and Elledge, 1996). CDKIs are expressed in response to a variety of anti-proliferative signals, including DNA damage and differentiation signals (Harper and Elledge, 1996). Three of the most prominent CDKIs (p16, p27 and p21) will be briefly described. p16 serves as a selective inhibitor of cyclin D1:CDK4, thereby acting to prevent phosphorylation and inactivation of pRB (Lee and Yang, 2001) (figure 3B). Cyclin E:CDK2 activity is repressed by the CDKI p27 (figure 3B). p27 accumulates to high levels in quiescent cells and must be inactivated for cells to enter S phase. In turn, inactivation of p27 is catalysed by cyclin E:CDK2, thereby ensuring fine control of cell cycle re-entry (Lee and Yang, 2001). Induction of p21 leads to inactivation of multiple cyclin/cdk complexes and allows pRB to remain hypo-phosphorylated. Hypo-phosphorylated pRB is capable of binding and inactivating of E2F causing cell to exit the cell cycle. pRB also activates tissue specific transcription factors which in turn up regulate p21 resulting in a regulatory loop. Studies on the expression of p21 found that the transcription factor HBP1 has a repressive effect on the p21 promoter (Gartel et al., 1998). It is unclear at the moment, to what extent the regulation of p21 by HBP1 is biologically relevant. Down regulating the cdk-inhibitory protein p21 has been linked to G1 progression while over expressing of HBP1 in a transgenic mice model leads to a delayed G1 phase of the cell cycle (Shih et al., 2001).

Retinoblastoma

Products of the mammalian retinoblastoma (Rb) tumour suppressor gene family are important regulators of cell cycle, differentiation and apoptosis. Studies of retinoblastoma patients led to the current Knudson "two- hit" hypothesis for tumourigenesis (Knudson, 1971). The first tumour-suppressor gene to be cloned was the Rb gene. In familial cases of retinoblastoma, a germline mutation in the Rb gene is inherited and affected individuals develop retinal tumours in which the loss of the second Rb allele is the rate-limiting step (Weinberg, 1991).

Human pRB is a nuclear phosphoprotein with a relatively long half- life (>8hrs). pRB is synthesized throughout the cell cycle and its activity is regulated by cell cycle-dependent phosphorylation. pRB is broadly expressed, but the levels of the protein are variable between cell types. pRB is a member of a family of three closely related mammalian proteins that includes p107 and p130 (Mulligan and Jacks, 1998). Together, these proteins are referred to as the 'pocket proteins' because their main sequence similarity resides in a domain (the pocket domain) that mediates interactions with viral proteins (Mulligan and Jacks, 1998). These interactions are important for the transforming properties of viral products and are thought to inactivate pRB. Although there are several ways that protein association might disrupt pRB function, the most widely held view is that either the viral proteins prevent pRB from interacting with its normal partners by directly competing for binding sites on pRB, or inducing pRB degradation (Dyson, 1994). Mutagenesis of several viral proteins demonstrated that a small conserved domain within these proteins, containing the amino acids LXCXE, is essential for the binding of these proteins to pRB. In contrast,

a relatively large portion of pRB, the so-called pocket domain, is required for this interaction and is essential for the growth suppression functions of pRB. It is indicative that most of the naturally occurring, tumour-derived mutations in Rb affect this region. However, regions outside of the pRB pocket are also likely to be important for its interaction with its partners (Dyson, 1994). The Rb family appears to be able to regulate cell cycle exit through at least two distinct mechanisms. One mechanism was described above; this involved repression of E2F transcriptional activity. The second mechanism is unique to differentiation process. During the differentiation process, pRB interacts and augments the activity of transcription factors necessary for differentiation. One of the identified interacting proteins is HBP1 and this interaction will be discussed in more detail below.

During differentiation in muscles (muscle cells represent the best characterized *in vitro* differentiation system), Rb and p130 expression is markedly increased. During differentiation of C2C12 myogenic cell line HBP1 is induced with approximately the same kinetics as pRB and p130. Forced over expression of HBP1 in the presence of pRB and p130 leads to cell cycle arrest suggesting that alongside pRB HBP1 might play a role in initiation and establishment of cell cycle exit by repressing transcription of cell cycle genes (see discussion on HBP1). Expression of HBP1 not only causes a cell cycle exit but also blocks muscle differentiation of the C2C12 cell line (Shih et al., 1998) Myf5, the first of the MyoD family of transcription factors expressed in differentiation muscle cells, up regulation is not affected by the forced HBP1 expression. HBP1 disrupts the transcriptional activity of Myf5 in a manner that is dependent on pRB or p139. This prevents the expression of later members of the MyoD family, under which myogenin and MyoD itself. This differentiation block can

be overcome by expression of the muscle specific transcription factors myogenin or MyoD. Furthermore, over expression of pRB also release the cells from the HBP1 imposed differentiation bock. The authors suggested that the relative ratio between pRB and HBP1 determines whether only cell cycle exit or full differentiation occurs. At low ratios of pRB/HBP1, cell cycle exit occurs but pRB mediated expression of muscle specific genes is blocked. When the ratio between pRB and HBP1 increases MyoD family is activated and full differentiation occurs. M. Lipinski and T Jacks (Lipinski and Jacks, 1999 156) suggested a possible model that integrates E2F and HBP1 regulation by pRB. During the early stages of differentiation, hypo-phosphorylated pRB accumulates. The hypo-phosphorylated pRB binds and inactivates E2F family factors, leading to decreased transcription of cell cycle genes. As HBP1 protein levels also increase, HBP1/pRB complexes can bind to and further inactivate promoters of these genes. This complementary action of E2F/pRB and HBP1/pRB complexes eventually would lead to transcriptional repression of mitotic genes and cell cycle exit followed by differentiation. Besides, in muscle differentiation HBP1 may be actively involved in the differentiation process during haematopoiesis. In these experiments, induction of maturation in myeloid cell lines (HL-60 and K562) resulted in an initial rapid decline in the level of HBP1 RNA, followed by a marked (20 fold) increase upon final maturation (Lin et al., 2001a). These results are compatible with the model described above and might indicate that different cell types utilize HBP1 and pRB to achieve a post mitotic state and full differentiation.

In the next chapter, HBP1 will be described in more detail. These descriptions include the structural properties, regulation of expression and the known target genes of HBP1.

HMG box containing protein 1

HBP1 is an ubiquitously expressed protein (Zhuma et al., 1999) that can potentially interact with retinoblastoma through two motifs (LXCXE and IXCXE) which are high and low binding sites respectively, for the pocket domain of retinoblastoma and p130 (RB1 and RB2 in figure 5A). These interactions were confirmed by co-immunoprecipitation of HBP1 with retinoblastoma in experiments where these proteins were co-expressed in cells by transient transfection (Lavender et al., 1997; Tevosian et al., 1997). HBP1 contains a masked activation domain. Deletion of two independent N- and C-terminal inhibitor domains unmasks the activation. The released activation capacity is repressed by retinoblastoma or its family members (Lavender et al., 1997).

Structural domains of HBP1

HBP1 is 513 amino acids long and contains two distinguishable structural domains, an HMG box and an Ataxin domain called AXH (figure 5).

The High Mobility Group

High Mobility Group (HMG) proteins were identified as a class of non-histone chromosomal species, which share the common property of rapid migration in SDS-PAGE gels. Represented in all four eukaryotic kingdoms and ubiquitous in mammalian cells, HMG proteins were first isolated from calf thymus chromatin and characterized in 1973 (Goodwin and Johns, 1973). Numerous HMG box proteins have since been identified. An evolutionary study of the HMG box proteins has led to the notion that two types of HMG box proteins can be distinguished (Soullier et al., 1999).

The first type usually contains multiple HMG boxes that bind non-specifically to DNA, recognizing distinct structural features such as pre-bent or cruciform DNA rather than specific sequences. These proteins are relatively abundant in all cells, on average one molecule per 10 to 15 nucleosomes in the nucleus (Soullier et al., 1999). These non-histone proteins probably play both a structural and a functional role in maintaining the overall chromatin structure.

The second type of protein contains a single HMG box that can bind to DNA in a sequence specific fashion (Soullier et al., 1999). This part of the HMG box family contains transcription factors which bind to specific DNA sequences, such as the T-cell transcription factors TCF1 and LEF1, the mammalian male sex determining factor SRY, numerous SOX factors and HBP1 (Soullier et al., 1999). Circular permutation assays indicated that these proteins can induce a strong bend of approximately 70-130° in the DNA helix (Bewley et al., 1998).

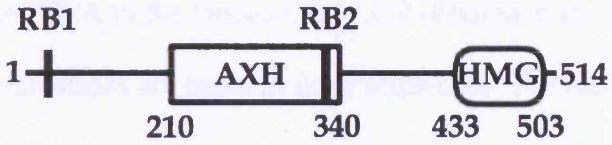
Figure 5: Schematic representation of the architecture of HBP1 and ATX1

A) The positions of the high and low affinity retinoblastoma binding sites (RB1 and RB2 respectively) and of the HMG and AXH motifs are indicated for HBP1.

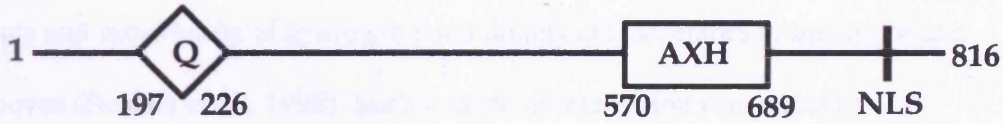
B) The polyQ tract, the nuclear localization signal (NLS) and the AXH module are marked for ATX1.

C) Multiple alignment of the AXH sequences of HBP1 and ATX1 families and two hypothetical proteins from C. elegans and D. melanogaster. Secondary structure prediction for the multiple alignment as obtained by the PHD programme is reported in the first line (E = β sheet, H= α helix).

A: HBP1



B: ATX1



C: Alignment AXH of HBP1 and ATX1

```
                    EEE  EEEEEE        HHHHHHHHHHH            EEEEE
HBP1_Human PSTVWHCFLKGTRLCFHKGSNKEWQDV--EDFARAEGCDNEEDLQMGIHKGYGSDGLKLLSHEESVSF  66
  HBP1_Rat PSTVWHCFLKGTRLCFHKESKKEWQDV--EDFARAATCD-EEEIQMGTHKGYGSDGLKLLSHEESVSF  65
HBP1_Mouse PSTIWHCFLKGTRLCFHKESNKEWQDV--EDFARAASCDNEEEIQMGTHKGYGSDGLKLLSHEESVSF  66
HBP1_Xlaevis PQTVWHCFLKGTRLCFHKGRKQWQDV--EDFAKSTRCRNERGIDSGTYKDYGSDGLKLVSYEECVSY  66
ATX1_Human PPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIDSSTVERIEDS-----------H  56
ATX1_Mouse SPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIHSSTVERIEES-----------H  56
ATX1_Rat SPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIDSSTVERIEDS-----------H  56
ATX1_Celegans STYYPTHFMRGTQLNVANGNIKKVEDLSDDFLRCAAESDDVIVNASVIKSIKS-------------  54
ATX1_Drome SDDSASCFRRGSYIELASGAMRRVEDIRTEDFIQSALRSQLFELREATVVRIDWS-----------G  56
     ruler 570.......580.......590.......600.......610.......620.............

                EEEEEEE        EEEEEEEE   EEEEE   EEE   EEEEEEE  EEEEEEEEEEEE
HBP1_Human GESVLKLTFDPGTXEDGLLTVECKLDHPFYVKNKVWSSFYPSLTVVQHGIPCCEVHIGNVCLPPGHP  133
  HBP1_Rat GESVLKLTFDPGTVEDGLLTVECKLDHPFYVKNKGWSSFYPSLIVVQHGIPCCEIHIGDVCLPPGHP  132
HBP1_Mouse GESVLKLTFDPGTVEDGLLTVECKLDHPFYVKNKGWSSFYPSLIVVQHGIPCCEIHIGDVCLPPGHP  133
HBP1_Xlaevis GESVLQLTFDPGTTDVGLLTVECKLDHPYYVKNKGWSSFYPSLIVVQHGIPCCEIHVGDVCLPPGHP  133
ATX1_Human SPGVAVIQFAVG-EHRAQVSVEVLVEYPFFVFGQGWSSCCPERISQLFDLPCSKLSVGDVCISLTLK  122
ATX1_Mouse SPGVAVIQFAVG-EHRAQVSVEVLVEYPFFVFGQGWSSCCPERISQLFDLPCSKLSVGDVCISLTLK  122
ATX1_Rat SPGVAVIQFAVG-EHRAQVSVEVLVEYPFFVFGQGWSSCCPERISQLFDLPCSKLSVGDVCISLTLK  122
ATX1_Celegans TAGSVTIIFETG-IEKQLIPLKCQVEHPFFVLGKGWCSCNPRKSGENYGLDCEILQVDDVCIVL-TR  119
ATX1_Drome CPSLVTLTFSYD-THHAKMDLQVQPGHPMFVYGQGWASCDPQLSLQLYELKCQQLQVGDICLSL-VP  121
     ruler ....630.......640.......650.......660.......670.......680.........
```

Compared to classical transcription factors the sequence specificity of the HMG box containing protein is fairly low (Weiss, 2001). For instance, HBP1 is able to bind AGAATGGG, TCAATGGG and AAAATGGG in the N-MYC promoter (Tevosian et al., 1997) and TTCATTCATTCA in the human CD2 LCR (Zhuma et al., 1999), but binding is inhibited when mutations are made in these sequences. The fact that these proteins seem "not precise" in recognizing a consensus sequence should not come as a surprise. It is remarkable that they can recognize a specific sequence at all. The interaction of sequence specific HMG boxes with linear DNA has been shown to proceed mainly through the minor groove, as revealed by methylation interference experiments and substitution of hydrogen bond donors and acceptors in the major and minor grooves (Bewley et al., 1998). Such a mode of interaction is unusual for sequence specific DNA binding proteins and is probably responsible for the modest specificity of HMG boxes for their binding sites. The minor groove provides little opportunity for base specific contacts, since hydrogen bonding cannot distinguish AT pairs from TA pairs. However, hydrogen bonding in the minor groove appears well suited for structure-directed recognition, since the phosphates will be spaced at favourable distances for selective interactions.

The AXH domain

A small motif (ca. 130 amino acids long), named AXH module, was identified in HBP1 and in the apparently unrelated transcription factor ATX1 whose aberrant regulation leads to spinocerebellar ataxia type-1 (SCA1) (Mushegian et al., 1997) (figure 5B). The AXH module is highly conserved within members of each of the two

protein species sharing ca. 78-80% identity. However, between the ATX1 and the HBP1 the identity is much lower (~28%) but still showing ~54% similarity (figure 5C).

The AXH of HBP1 coincides with the repression domain of HBP1. The repression activity is intrinsic to the domain but can be modulated by Rb and p130 (Tevosian et al., 1997).(Yue et al., 2001) Besides the interaction with Rb and p130 the AXH domain can also interact with Sin3 (Swanson et al., 2004) and TCF4 (Sampson et al., 2001). SIN3 is a multi-protein complex that represses the transcription of many eukaryotic genes. The SIN3 complex does not directly bind DNA but is targeted to specific genes through protein:protein interactions between SIN3 and DNA-binding proteins in this case HBP1 (Swanson et al., 2004). The mammalian SIN3 complex contains chromatin-modifying enzymes such as HDACs. The interaction between TCF and AXH can suppress the Wnt signalling (Sampson et al., 2001) and this interaction will be discussed further on.

The AXH of ATX1 partially overlaps with a protein interaction domain, causing dimerization of the protein (Burright et al., 1997) and is contained in a region involved in RNA binding. No ATX1 gene targets haven yet been reported.

Controlling HBP1 levels

The concentrations of transcription factors, as of intracellular proteins in general, may be regulated at any of the steps leading from DNA to protein, including transcription,

RNA processing, mRNA degradation and translation. The activity of a transcription factor is often regulated by phosphorylation, which may affect different functions. In the following paragraph, an overview of different modes of regulation of HBP1 will be described.

The promoter activity of HBP1 has been studied in a promyelocytic cell line (Austin and Lin, 2002; Lin et al., 2001b). In these cells, the proximal 1.1 kb segment of the 5' flanking region of the HBP1 gene exhibited strong promoter activity. Six hours after induction of differentiation, the promoter activity decreased to almost zero. This is followed by a dramatic increase of up to 50 times the pre-treatment levels after 24 hours (Austin and Lin, 2002; Lin et al., 2001b).

The dramatic increase in HBP1 RNA levels, which accompanies myeloid differentiation, is due to an increase in HBP1 promoter activity. Transcription of the HBP1 gene appears to be regulated by AP1 and c-myb, which enhance HBP1 promoter activity, whereas GM-CSF and/or PEA3 inhibit HBP1 promoter activity (Austin and Lin, 2002; Lin et al., 2001b).

Naturally occurring antisense transcripts have been documented in prokaryotes and viruses, where they are found to regulate gene expression by affecting mRNA transcription, processing and translation. A growing number of endogenous antisense RNA transcripts have also been reported during the last several years in a variety of eukaryotic organisms. The effects that these eukaryotic antisense RNAs have on the corresponding sense RNAs has not been clearly established (Good, 2003).

Recently, Rodrigo Yelin et al.(Yelin et al., 2003) identified transcripts that derive from both strands of the HBP1 locus. Thus, it is possible that antisense mediated gene regulation plays a role in HBP1 expression.

HBP1 is also regulated post-transcriptionally. It was found that p38 MAP kinase both binds and specifically phosphorylates HBP1 at serine 401(Xiu et al., 2003). As a result, HBP1 stability is increased. Only recently, the p38 MAP kinase pathway has been implied in the regulation of the cell cycle. p38 has to be inactivated to exit from the quiescent state as well as to elude mitotic arrest. Moreover, activation of p38 MAP kinases during the G1/S transition can either inhibit or promote cell cycle progression depending on the stimulus and the cell type (Ambrosino and Nebreda, 2001). Thus, increasing the levels of HBP1 by p38 MAP kinase may contribute to the cell cycle regulation (figure 6).

Targets of HBP1

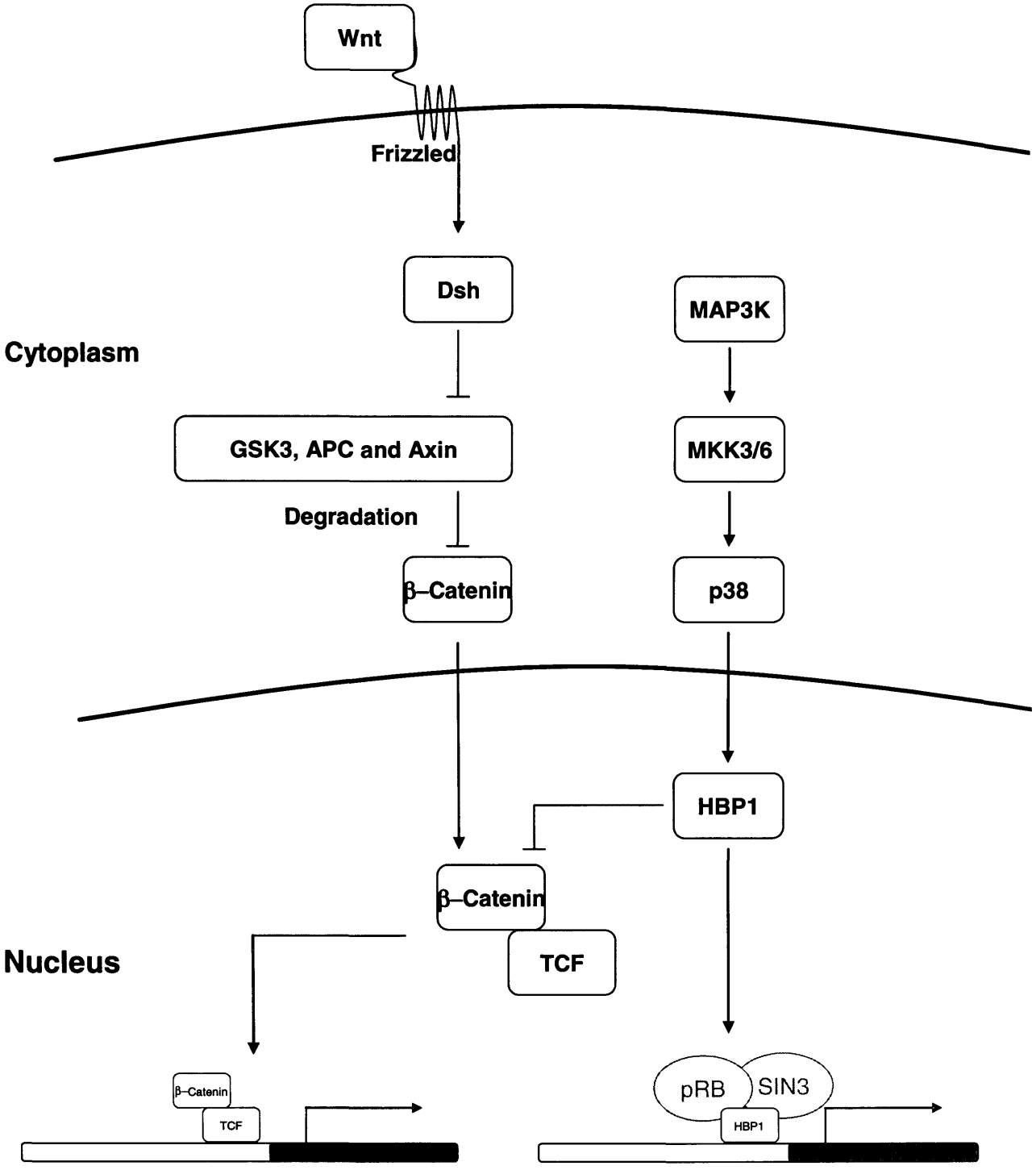
HBP1 encodes a transcription factor that is up regulated during differentiation and regulates a specific set of target-genes. In the following chapter a brief description of the some of the target genes that of HBP1 will be given.

Figure 6: Overview of HBP1 regulation and its targets

Levels of HBP1 proteins in the cell are post-translational regulated by p38 phosphorylation. Phosphorylation by the p38 MAPK increases the protein stability of HBP1 resulting in increased protein concentration.

HBP1 can repress the Wnt signalling pathway. Binding of Wnt ligands to their cell surface receptors Frizzled activates cytoplasmic Dishevelled (Dsh) which interferes with the function of the destruction complex consisting of GSK-3, APC, and Axin. This prevents β-catenin degradation. As a consequence, β-catenin is allowed to enter the nucleus where it binds to TCF transcription factors and functions as a transcriptional activator to induce expression of specific Wnt target genes. HBP1 can compete with β-catenin for binding to TCF transcription factors and inhibits the expression of specific Wnt target genes.

HBP1 can bind to the regulatory elements of target genes and can either activate or repress transcription. It can influence expression levels directly or indirectly through the binding to pRB and/or the SIN3 complex.



Cyclin D1 gene

C-MYC

Down regulation:

- **N-myc**
- **p21**

Up regulation:

- **hCD2**
- **Histone $H1^0$**
- **MPO**

Wnt signalling

Several proteins, starting with interaction of Wnt with its receptor Frizzled (Fzd), participate in the Wnt signalling cascade (Novak and Dedhar, 1999). Frizzled interacts with and activates Dishevelled (Dsh) leading to the down-regulation of the activity of the GSK3β and the stabilization of cytoplasmic β-catenin (figure 4) (Novak and Dedhar, 1999). Cytoplasmic β-catenin becomes available for active transport to the nucleus by an importin-β-dependent mechanism (Novak and Dedhar, 1999). There β-catenin participates in transcription regulation through its interaction with the TCFs (Brantjes et al., 2002). The latter constitute a family of HMG box containing transcription factors originally identified for their ability to bind the T cell receptor enhancer (Brantjes et al., 2002). Expression of TCFs (TCF1, LEF1, TCF3 and TCF4) is not restricted to lymphoid tissue and they have now been implicated in a variety of developmental processes (Brantjes et al., 2002) and in cancer (Giles et al., 2003). A functional TCF transcription complex requires the assembly of the TCF protein with β-catenin. β-catenin provides the transcriptional activation domains while TCF provides the DNA binding domain (Brantjes et al., 2002). Several known transcriptional targets get up regulated through binding of TCF:β-catenin complex to their promoters, such as c-MYC and Cyclin D1 whose promoters contain TCF binding sites and appear to be constitutively active in several cancers (Giles et al., 2003).

HBP1 can suppress the Wnt signalling pathway (Sampson et al., 2001). It was shown that both Cyclin D1 and c-MYC expression was extinguished upon HBP1 over expression. HBP1 targets the transcriptional activators TCF4, which results in the loss of activity of TCF4. HBP1-DNA binding is not required for inhibition of Wnt-ß-catenin transcriptional activation. Within HBP1, the necessary functional region for interacting with TCF4 is defined to amino acids 192-400, which is contained within the, previously described, AXH module (Sampson et al., 2001). While DNA binding was not necessary for down regulation of Cyclin D1 and c-MYC, these promoters do contain weak HBP1 binding sites (Yee et al., 2004). Further studies will be required to address what the effect is of direct binding to these promoters by HBP1(Tevosian et al., 1997).

N-Myc

The transcription factors encoded by the myc genes are implicated in the control of several aspects of cellular differentiation including cell proliferation and apoptosis (Grandori et al., 2000). Lack of N-myc expression causes severe effects of embryonic development while over expression of N-myc is associated with the development of several types of childhood tumours. HBP1 repressed the rat N-myc promoter in transient assays, it was speculated that HBP1 and E2F binding sites allow for the integration of opposing signals at the N-myc promoter via HBP1 or E2F in conjunction with pRB (Tevosian et al., 1997). There is no evidence yet to support this regulation of endogenous N-myc by HBP1.

Histone $H1^0$

The replacement linker histone $H1^0$ is associated with terminal differentiation in many mammalian cell types, and its accumulation in chromatin may contribute to transcriptional repression occurring during terminal differentiation. HBP1 was identified as a target of histone $H1^0$ in a yeast one hybrid using the element that regulates $H1^0$ expression pattern as bait (Lemercier et al., 2000). HBP1 can up regulate histone $H1^0$. Furthermore, it was shown that Rb could stimulate the DNA binding activity of HBP1 to the histone $H1^0$ binding site. Through modulating the expression of specific chromatin-associated proteins such as $H1^0$, HBP1 can play a vital role in chromatin remodelling events during arrest of cell proliferation in differentiating cells.

Myeloperoxidase

Myeloperoxidase (MPO) is a human enzyme in the azurophilic granules of neutrophils and in the lysosomes of monocytes. Its major role is to aid in microbial killing. HBP1 binds to Myeloperoxidase promoter and enhances MPO promoter activity (Lin et al., 2001a). However, high levels of HBP1 expression are not associated with high levels of Myeloperoxidase mRNA. At the moment, it is unclear what the cause of this discrepancy is.

Summary of HBP1

In summary, HBP1 appears to be an important regulator in the cell cycle and differentiation. Studies have shown that HBP1 can act as both a transcriptional activator and repressor. In vivo, it was shown that transgenic mice expressing HBP1 had a delayed G1 phase of the cell cycle (Shih et al., 2001). Possible mechanisms through which HBP1 can repress the cell cycle was suggested by experiments which showed that HBP1 can repress the N-myc (Tevosian et al., 1997) and cyclin D1. Specific gene activation of N-myc and cyclin D1 have been identified as the cause of some cases of leukaemia (de Boer et al., 1997; Mirro, 1992). The stoichiometry of Rb and HBP1 contributes to differentiation (Shih et al., 1998). HBP1 up regulates histone $H1^0$, a linker histone variant that is expressed in terminally differentiated cells concomitant with the arrest of cell proliferation.

Gene targeting

Embryonic mouse stem cells

The ability to generate a mouse with a targeted mutation in a desired gene has been one of the most important technological advances in our efforts to understand the function of gene products. Not only can gene disruption demonstrate the function of genes, but by disrupting development or survival of particular cells, it can also

demonstrate the function of particular cells in the whole animal. Some knockout mice proved to be murine models of human genetic diseases and have shown that a single gene defect is capable of causing disease. In genetic terms, targeted gene disruption is complementary to using mutant strains of mice with a known phenotype to map and identify the disease causing genes.

Gene targeting is the term used to describe the predetermined mutation of an endogenous gene that results from a homologous recombination event between a mutated version of the gene carried on a targeting vector and the endogenous locus. When performed in embryonic stem (ES) cells, gene targeting is a powerful means for generating mice carrying specific genetic mutations.

Homologous recombination between the targeting vector and the endogenous locus in ES cells is a rare event, but gene targeting vectors incorporate marker genes for drug selection schemes to enrich for the recovery of homologous recombinant ES cells. Gene targeting experiments can employ different strategies to obtain a functional knockout. Basically, targeting vectors can disrupt gene expression by deleting certain sequences and replacing them with substitute DNA, usually a drug selection gene or a reporter gene to analyze target gene expression (Thomas and Capecchi, 1987). Alternatively, targeting vectors can introduce loxP sites for inducible or tissue specific deletions of the target gene or they can introduce new gene segments without deleting parts of the endogenous locus.

Cre LoxP system

Conventional gene knockout technology by homologous recombination can provide important information toward elucidating the function of genes; however, the role of many genes cannot be investigated due to early embryonic lethality. Alternatively, the role of a particular gene in adult tissues may be masked by developmental abnormalities in any knockout animal generated. It is also possible that a phenotype can only be seen shortly after deletion, before compensatory mechanisms are activated, and take over the function of the inactivated gene. The implementation of site-specific recombinases, such as the bacteriophage P1 loxP/CRE system, enables the development of conditional knockouts that lack a particular gene or domain of a gene only in a specific tissue or after a specific stage of development. This system can also be used to facilitate transgene activation/inactivation in vivo, deletions of large stretches of chromosomal DNA, chromosome translocations and subtle alterations to genes and/or their regulatory elements.

The CRE enzyme is a 35-kDa protein isolated from bacteriophage P1, which acts as a site specific recombinase (Rossant and Nagy, 1995). The sequence recognized by CRE, termed loxP, consists of two 13 base pair inverted repeats (CRE recognition sites) separated by an 8bp spacer. The 8-bp spacer region is asymmetrical and is responsible for the directional nature of the LoxP site. Two loxP sequences in the same orientation will result in excision of the intervening DNA by CRE leaving one loxP site. In contrast, where two loxP sites are in opposite orientation, CRE will invert the intervening piece of DNA. This method can also be adopted to translocate, delete, or invert large pieces of DNA only where and when required by the investigator. The

latter is limited only by the availability of suitable regulators of CRE recombinase expression.

The success of the CRE/LoxP system relies on the targeted gene to function normally until recombination occurs. As mentioned earlier resistance genes are included in the targeting vector. Resistance genes included for selection of correctly targeted clones usually include ≥ 1kb of foreign DNA, which has the potential to influence the expression of the targeted gene prior to recombination (Meyers et al., 1998; Nagy et al., 1998). There is also some evidence that resistance genes located in introns can be spliced to upstream exons, effectively generating a truncated or mutant protein. To minimize this potential influence on gene expression, it is necessary to splice out the resistance gene from the correctly targeted embryonic stem cell. To remove the resistance gene, an additional loxP must be included to flank the resistance gene. The modified genomic DNA would thus include three loxP sites, two flanking the resistance gene and a third located so that it can cause the desired mutation when recombination occurs.

Rationale

The aim of this study is to determine the function of HBP1 in development and physiological processes in vivo. Structure provides the link between sequence and function, and therefore is central to unravelling the function of HBP1. HBP1 contains two recognisable domains: The well-characterised HMG box and a novel domain called AXH. We therefore set out to define the structural properties of this novel domain in collaboration with the department of Molecular Structure here at the National Institute for Medical Research.

A significant limitation of the structural approach is that it downplays the role of context (i.e. when, where and how the genetic information is utilized) in protein function. It would therefore be impossible and at times even misleading to predict its function, because a purely structural description ignores the context in which proteins operate. Initial analysis was performed to define the subcellular distribution of the HBP1 protein in order to determine how HBP1 behaves in the context of a single cell. Furthermore, these subcellular distribution studies identified which regions of the HBP1 protein are necessary for the correct distribution within the cell.

In complex organisms, gene function is defined by the temporal, developmental and physiological patterns in which a gene is expressed. In these organisms, tissue or whole animal context alone determines the final functionality, and all types of regulation of the activity of protein (gene expression, subcellular localization, post-

translational modifications) are functional attributes. Therefore, we set out to generate mice in which HBP1 is inactivated. Studying the development and physiology of such mice would allow us to determine the function of the HBP1 gene in the context of the whole organism.

HBP1 is a ubiquitously expressed protein and given that there are no close homologues that could replace its function in its absence, there was the possibility that mice deficient in HBP1 may be embryonic lethal. To avoid this, it would be desirable to delete the gene in a tissue specific manner. The ability to inactivate an endogenous gene in the mouse in a temporally and spatially controlled manner is not only useful for circumventing early lethal phenotypes, but also allows biological questions to be addressed with exquisite accuracy. To achieve this, an essential region of HBP1 was flanked by loxP sites (floxed), so that this domain could be deleted specifically at will by directing the expression of Cre in a tissue specific manner. For this purpose, expression of the Cre recombinase was directed in the haematopoietic compartment. In order to achieve Cre expression in the entire haematopoietic compartment or only in lymphocytes, we have chosen to express Cre under the control of VAV regulatory elements and the human CD2 regulatory elements, respectively.

In order to generate mice in which an essential region of the HBP1 gene is floxed a targeting vector was transfected into mouse ES cells. LoxP sites and drug resistance markers were assigned to introns, so that their location would not affect transcription of the targeted allele. As drug resistance gene insertion sometimes affects the expression of the target gene this unit was also flanked by loxP sites. Thus, the targeting vector has three loxP sites, positioned such that neo and the essential region

of the gene of interest were floxed. In this case, a partial Cre mediated recombination event would result in genomic deletions, one of which removes neo, leaving the essential region floxed.

Materials and Methods

General

Chemicals and reagents

Restriction endonucleases, T4 DNA ligase, Klenow DNA polymerase, Proteinase K, DNase I, alkaline phosphatase and RNase A were obtained from Boehringer Mannheim and New England Biolabs. Hybond nylon membranes were purchased from Amersham. Chemicals were obtained from BDH, Sigma Chemical Company and Pharmacia Biotech. Antibodies were purchased from Pharmingen and Caltech. Tissue culture plastics were obtained from Corning-Costar, Becton Dickinson and Nalgene Nunc International. Foetal calf serum (FCS) and other media supplements were purchased from Globepharm limited, GIBCO-BRL and Sigma Chemical Company.

Molecular biology

Bacteriological cultures

Liquid cultures were grown in L-Broth supplemented with 100μg/ml ampicillin. For agar plates, L-Broth was supplemented with 1.5% (w/v) bactoagar. Bacteria were frozen in 1 ml aliquots of L-Broth supplemented with 1/10 volume of 10x Hogness.

L-Broth:

1% (w/v) bactotryptone; 0.5% (w/v) yeast extract; 1% (w/v) NaCl in ddH_2O.

Hogness (10x):

36mM KH_2PO_4; 13mM K_2HPO_4; 20mM NaCitrate; 10mM $MgSO_4$; 40% (w/v) Glycerol.

Competent bacteria and cell transformation

For competent bacteria, a single colony of Escherichia coli (E.coli) DH5α bacteria was picked from agar plates and grown (250rpm in a New Branswicki Model G25 orbital shaker) overnight to stationary phase at 37°C in 2.5ml of L-Broth. 250ml of L-Broth with 20mM Magnesium sulphate was inoculated with 2.5 ml of the overnight culture and grown with shaking at 37°C until an optical density of 0.5 at 600nm was reached. Bacteria were then pelleted in 50ml falcon tubes (Falcon, Beckton Dickinson) at 1285g, 4°C for 20 minutes and gently resuspended in 100ml of ice cold TFB1. Cells were left on ice for 5 minutes. Bacteria were then pelleted at 1285g, 4°C for 20 minutes and then re-suspended in 10 ml of ice cold TFB2. Cells were incubated

for 45 minutes on ice and then frozen in 100µl aliquots in sterile Eppendorf tubes in a dry ice/methanol bath and stored at -70°C.

For transformation of bacteria, a ligation reaction containing 100ng of DNA was mixed with 100µl of competent cells. Following incubation on ice for 30 minutes, the cells were heated to 42°C for 30 seconds and placed back on ice for another 2 minutes. 1 ml of L-Broth was added and cells were grown for 45 minutes at 37°C. 100µl of this culture was then plated out onto L-agar plates supplemented with the correct antibiotics.

TFB1:

30mM potassium acetate, 10mM $CaCl_2$, 50mM $MnCl_2$, 100mM RbCl 15% glycerol. pH adjusted to 5.8 with 1M acetic acid.

TFB2:

10mM MOPS (pH 6.5), 75mM $CaCl_2$, 10mM RbCl 15% glycerol. pH adjusted to 6.5 with 1M KOH.

Small scale plasmid DNA isolation

1.5 ml of a 5ml overnight culture of L-Broth supplemented with antibiotics was transferred to an Eppendorf tube and pelleted by centrifugation at 13000xg for 1 minute. The bacterial pellet was resuspended in 200µl STET and 20 µl 10mg/ml lysozyme. The mixture was heated to 100°C for 1 minute. Debris was pelleted for 10 minutes at 13000xg in a bench top centrifuge. The pellet was removed and the plasmid DNA was precipitated from the supernatant with 600µl of 100% ethanol at

4°C and collected by centrifugation at 13000xg. The DNA pellet was resuspended in 100µl ddH_2O and stored at -20°C.

STET buffer:

0.1 M NaCl; 10 mM Tris-Cl, pH 8.0; 1 mM EDTA, pH 8.0; 5% Triton X-100

Large scale plasmid DNA isolation

Bacteria were harvested from 1litre over night culture by centrifugation at 4000rpm for 30 minutes, at 4°C. The pellet was resuspended by vortexing in 40ml 1x glucomix and incubated for 5 minutes at RT. Bacteria were then lysed with 80ml of Lysis Mix for 10 minutes and chromosomal DNA and protein was precipitated for 10 minutes on ice by the addition of 40ml 5M KOAc. Debris was removed by centrifugation for 20 minutes at 4000rpm. The supernatant was filtered through two layers of cheesecloth and plasmid DNA was precipitated at RT by the addition of 0.6 volumes of isopropanol and harvested by centrifugation for 20 minutes at 5000rpm. The pellet was dissolved in 9ml of ddH_2O. 10g of CsCl and 0.5ml of 10mg/ml ethidium bromide were added to the DNA solution. Samples were then transferred into Quickseal Ultracentrifuge Tubes (Beckman) and centrifuged at 55000 rpm for 15 hours in a Beckman VTi65.2 rotor. Banded plasmid DNA was recovered using a hypodermic needle and syringe and ethidium bromide was removed by repeated extraction using butanol saturated with ddH_2O. The DNA solution was then diluted with 3 volumes of ddH_2O and DNA was precipitated with 2.5 volumes of ethanol. After centrifugation at 8000rpm the DNA pellet was resuspended in 1ml ddH_2O. RNase A was added to the DNA to a final concentration of 20mg/ml and it was incubated at 37°C for 30 minutes. The DNA was then extracted once with equal volume of

phenol:chloroform:isoamylalcohol 25:24:1 and precipitated by adding 3M NaOAc to 0.3M final concentration and 2.5 volumes of ethanol. The DNA pellet was washed with 70% ethanol, resuspended in 1ml ddH_2O and stored at –20°C.

Glucomix (10x):

250mM Tris-HCl, pH8.0; 500mM Glucose; 100mM EDTA.

Lysis Mix:

200mM NaOH; 1% (w/v) SDS.

5M KOAc, pH4.8:

3M with respect to potassium; 5M to acetate; adjusted to pH4.8 with glacial acetic acid.

Lambda DNA preparation

A single colony of XL1 blue MRA strain was inoculated in 50 ml of LB broth supplemented with 0.2% of maltose and 10mM $MgSO_4$ and grown overnight shaking at 30°C. Bacteria were then pelleted in a sterile conical tube at 1285g for 10 minutes and gently resuspended in 10 mM $MgSO_4$ in the volume so that a OD_{600} of 0.5 was reached. 600µl of this mixture was combined with ~ $5x10^4$ pfu of bacteriophage. This mixture was incubated for 15 minutes at 37°C. 6.5 ml of melted (48°C) LB agar was added and spread evenly onto 150-mm bottom LB agar plate. The plate was incubated at 37°C for 6-8 hours. The plates were overlaid with 10 ml of SM buffer and incubated overnight at 4°C while rocking. The SM buffer was collected from the plates chloroform was added to a final concentration of 2%, vortexed and spun down

for 10 minutes to remove residual agarose. The lambda DNA was isolated using Qiagen lambda kit according to manufacturers' instructions.

SM (suspension medium)

50mM TRIS·Cl, pH8.0; 0.1M NaCl; 8mM $MgSO_4$; 0.01% gelatin

DNA isolation from PAC clones

A single colony was inoculated into 2 ml LB media supplemented with 25 μg/ml kanamycin. The cultures were grown overnight (up to 16 h) shaking at 225-300 rpm at 37°C. The bacteria were pelleted by centrifugation at 3500rpm for 10 minutes at 4°C. The bacterial pellet was then resuspended in 200μl 1x glucomix supplemented with Rnase, transferred to an Eppendorf tube and incubated at room temperature (RT) for 5 minutes. Bacteria were then lysed with 400μl of Lysis Mix and incubated on ice for 10 minutes. This was followed by the addition of 200μl 5x KOAc pH4.8 and gentle mixing. Debris was pelleted for 10 minutes at 13000rpm in a bench top centrifuge and plasmid DNA was precipitated from the supernatant with 0.6 volumes of isopropanol at RT and collected by centrifugation at 13000rpm. The DNA pellet was then resuspended in 100μl ddH_2O and stored at –20°C.

Genomic DNA preparation

Five mm of tail tissue was incubated at 55°C overnight in 500μl tail mix supplemented with 200μg/ml of proteinase K in a shaking water bath. The DNA was then extracted twice by shaking for 10 minutes with equal amount of

phenol:chloroform:isoamylalcohol 25:24:1 followed by centrifugation at 13000rpm in a bench top centrifuge for a further 10 minutes. This was followed by a similar extraction with chloroform. DNA was then precipitated with equal volume of isopropanol. The DNA was immediately spooled out using a Pasteur pipette which had previously been sealed in a Bunsen burner. The DNA was washed with 70% ethanol, air dried briefly and then resuspended in ddH_2O overnight at 4°C. DNA concentration was then determined by spectrophotometry.

Tail mix:

50mM Tris-HCl, pH8.0; 10mM EDTA; 1% (w/v) SDS.

Phenol:Chloroform:

Phenol:Chloroform:Isoamylalcohol (25:24:1 v/v) equilibrated with 10mM Tris-HCl, pH7.5.

DNA restriction digests

Restriction digests were carried out according to the manufacturers' directions. DNA was incubated under the appropriate conditions for the restriction enzyme used. For digests with more than one enzyme, which had not the same buffer requirements, DNA was digested in consecutive single digests. The DNA was phenol:chloroform extracted and ethanol precipitated between these digests. For enzymes that required the same restriction buffer, both enzymes were used simultaneously in the same reaction.

Agarose gel electrophoresis

Agarose gels were prepared in 1x TAE and run horizontally and submerged in 1x TAE buffer. This buffer was supplemented with 5mg/ml of ethidium bromide and run at 4V/cm. DNA was visualized and photographed under short wave UV light.

50x TAE:

2M Tris-Acetate; 100mM EDTA; pH8.0.

Extraction of DNA from agarose gels

DNA digests were run in agarose gels prepared in 1x TAE and bands were visualized under long wave UV light. The DNA was extracted from the gel using a GeneCleanII kit (Anachem) following the manufacturers instructions.

Southern blot

10µg of genomic DNA was digested in a final volume of 50µl overnight in the appropriate conditions for the restriction enzyme using 40 units of enzyme. To check completion of digestion, 3µl of this reaction was mixed with 0.5µg of lambda DNA in a second Eppendorf tube and incubated in parallel. The next morning the lambda DNA samples were run on a 0.8% agarose gel to determine the extend of digestion. Digests were then electrophoresed slowly (2V/cm) overnight in a 0.7% agarose gel

followed by photography under short wave UV light. The gel was then prepared for blotting by submerging it for 40 minutes in Blot I solution. The gel was then incubated for a further 50 minutes in Blot II solution. It was then set up for overnight blotting similar to Southern (Southern, 1975). DNA was then fixed onto the membrane using an optimized UV cross-linking protocol.

Blot I solution:

0.5M NaOH; 1.5M NaCl.

Blot II solution:

0.5M Tris-HCl, pH7.5; 3M NaCl.

RNA preparation

Total RNA was prepared from tissues. Total RNA isolation was carried out using RNeasy Maxi Kit (Qiagen) following the protocol as specified by the manufacturer. Integrity of RNA was determined by loading 10μg of RNA on a 1.2% agarose gel. The samples were electrophoresed at 10-15V/cm for 1 hour. Only samples with a 2:1 ratio between 28S and 18S ribosomal bands where used for downstream applications.

Northern Analysis

10μg of total RNA was loaded per lane on a 6% (v/v) formaldehyde gel and samples were electrophoresed at 5-7V/cm for 3.5 hours. The gel was then prepared for blotting by submerging it in double distilled H_2O for 2 times 15 minutes and in 20x SSC for

30 minutes. The RNA was transferred overnight on to a Hybond N+ nylon membrane similar to Southern (Southern, 1975). The RNA was fixed on to the membrane by using an optimized UV cross-linking protocol.

DNA probe labelling

DNA probes were radioactively labelled by random priming using a random priming kit (Pharmacia). 50ng of DNA resuspended in ddH_2O was denatured at 100°C for 5 minutes, cooled immediately on ice for 5 minutes and incubated in a final volume of 50µl with 5µl of [α32P]-dCTP (370MBq/ml, Amersham) and one Ready To Go™ DNA labelling bead (-dCTP) at 37°C for 15 minutes. Unincorporated nucleotides were removed using a Pharmacia 50G nick column according to the manufacturers' instructions. The collected probe was denatured by boiling at 100°C for 5 minutes before use in hybridisation reactions.

Filter hybridization

Filter membranes were pre-hybridised at 65°C for at least 30 minutes in 25ml of hybridisation buffer in either Hybaid bottles placed in a rotating Hybaid oven or in Perspex boxes submerged in a shaking water bath. 50ng of labelled probe was added to the hybridisation buffer and left to hybridise overnight at 65°C. Following hybridisation, nylon filters (Hybond XL, Amersham) were washed in a final

stringency of 0.1x SSC; 0.1% (w/v) SDS at 65°C changing wash solutions every 20 minutes. Filters were then prepared for autoradiography.

SSC (20x):

3M NaCl; 0.3M NaCitrate.

Hybridisation buffer:

10x Denhardts; 3x SSC; 0.1% (w/v) SDS; 10% (w/v) Dextran Sulphate; 50mg/ml denatured salmon sperm DNA.

100x Denhardts:

2% (w/v) BSA; 2% (w/v) Ficoll 400; 2% (w/v) Polyvinyl pyrollidone.

Autoradiography

Washed nylon membranes from hybridizations were wrapped in Saran Wrap and exposed to Kodak Biomax photographic film at -70°C using intensifying screens.

Membrane stripping

Radioactive probe was removed from hybridisation membranes by immersion of the membrane into a boiling solution of 0.1x SSC; 0.1% (w/v) SDS and letting the solution cool to RT. Complete removal of probe was confirmed by exposure of the membrane to photographic film.

Mouse genomic PAC library RPCI21 screening

The mouse PAC library RPCI21 provided by the HGMP Resource Centre was constructed by Kazutoyo Osoegawa working in Pieter de Jongs' laboratory at the Roswell Park Cancer Institute, Buffalo (Osoegawa et al., 2000). Filters were hybridized with (32)P-labelled (random-primed) probe ($1x10^6$ cpm/ml hybridization solution) using standard conditions. After washing to remove unbound probe, filters are wrapped in plastic film (Saran) and exposed to X-ray film (Kodak Biomax) for 24 hrs. Complete documentation on interpretation of hybridization results (location of positive clone plate coordinates) was included with the filter sets.

Sequencing

50 ng of purified plasmid DNA was sequenced with 3.2 pmoles of the appropriate primers and 8 µl of Perkin Elmer ABI Prism Big Dye Terminator cycling reaction in a final volume of 20 µl according to manufacturer's instructions. Cycle sequencing was carried out using a Hybaid MBS PCR machine with the following parameters: 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes for 25 cycles. Extension products were precipitated and simultaneously purified from unincorporated Dye-terminators by ethanol precipitation at room temperature for 15 minutes followed by centrifugation at 13.000xg for 20 minutes. Pelleted extension products were washed

with 70% ethanol and spun for 5 minutes. The pellets were air dried and resuspended in 5µl of loading buffer.

Sequence data was acquired on an Applied Biosystems (ABI) Model 377 Automated DNA sequencer with 36 cm well to read plates and 5% poly-acrylamide gels using standard running conditions with a 9 hour collection time. Sequence data was analyzed and edited with ABI Sequence Analysis. Edited DNA sequence was further analyzed using a Contig Assembly Program from www.tigem.it, Redasoft visual cloning and Blast software form NCBI.

5% Polyacrylamide gel:

5 ml 10x TBE; 5ml Long ranger acrylamide solution, FMC Bioproducts; 18 g of urea; ddH_2O to 50ml; 250µl of 10% (w/v) APS; 25µl TEMED

Loading buffer:

5 parts deionized formamide; 1 part 25 mM EDTA (pH 8.0) containing 50 mg/ml Blue dextran.

Genotyping by PCR

100ng DNA were added to a mixture containing: 1x Taq Polymerase buffer (Perkin-Elmer), 20nM dNTP's (Amersham), 2mM $MgCl_2$, 40pM appropriate primers, 3 units Taq polymerase (Perkin-Elmer) in a final volume of 20µl. The PCR was carried out in a Hybaid PCR machine and the products analysed on an agarose gel.

Primers:

P1 TTAATATGGTGGTGGCTTGATT

P2 GGCCAGCCAAAATATAATGC

P3 TAGCTGAAGAATGCCCCAAG

P4 CAGAAGCCTGCAAGGATGTT

PCR conditions:

94°C 1 min.

Then for 35 cycles: 94°C for 30 sec

56°C for 45 sec

72°C for 45 sec

72°C 2 min

4°C hold

Cellular biology

Expansion and trypsinisation of ES cells

PC3 (O'Gorman et al., 1997) embryonic stem cells from 129/Sv mice were maintained in the undifferentiated state by growth on a feeder layer of γ-irradiated primary embryonic fibroblasts using culture medium supplemented with leukaemia inhibitory factor (LIF) (Mansour, 1988). The ES cells were grown to confluency on 6cm culture

dishes in 5ml of ES medium. The culture dishes were pre-coated with 3 ml of 0.1% Gelatine for 30 minutes and contained a monolayer of $2x10^6$ embryonic fibroblasts. The culture medium was changed every day and when the cells reached confluency, they were washed with 5 ml of PBS and 0.5 ml of trypsin was added to them. The cells were put in 37°C for 3 minutes and they were pipetted up and down vigorously with a Pasteur pipette ensuring that a single cell suspension was obtained. The cells were spun out through medium, resuspended in 5 ml of ES medium and 1/10 of them were transferred to a fresh 6 cm culture dish with a feeder layer.

ES Medium:

450ml Dulbecco's Modified Eagle's Medium (DMEM) with 0.11G/L NaPyruvate, pyridoxine; 75ml Foetal calf serum (FCS); 5ml Glutamine 100x stock at 200mM; 2.5ml Penicillin/ Streptomycin 200x stock at 10^4 units penicillin and 10mg/ml streptomycin; 5ml Non essential amino acid (MEM) 100x stock; 3.5µl β-mercaptoethanol; 50µl Leukaemia inhibitory factor (LIF) 10^7 units/ml.

Phosphate buffered saline (PBS):

1.4M NaCl; 27mM KCl; 81mM Na_2HPO_4; 15mM KH_2PO_4; adjusted to pH 7.5 with HCL

Trypsin:

To 1 litre of ddH_2O add: 2.5gr Trypsin powder; 0.4gr EDTA; 7.0gr NaCl; 0.3gr Na_2HPO_4. $12H_2O$ or 0.22gr Na_2HPO_4. $7H_2O$; 0.24gr KH_2PO_4; 0.37gr KCl; 1.0gr D-glucose; 3.0gr Tris; 1ml Phenol Red 0.5%; adjusted to pH7.6 and filtered.

HEK-293 cells

The eukaryotic cells HEK293 were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % foetal bovine serum (FBS), supplemented with glutamine. Cells were incubated at 37°C in a 5 % CO2 atmosphere and passaged when cells reached confluence of 70-80 %.

Preparation of embryonic fibroblasts

To make G418 resistant embryonic fibroblasts (EFneo), a β2m-/- or an IL3-/- male mouse was mated to a wild-type female. The pregnant female was sacrificed at day 13 or 14 of gestation and the embryos were dissected out free of extra-embryonic membranes in sterile conditions. Each embryo was washed in 70% ethanol and subsequently in PBS and soft organs and viscera (liver, gut, and brain) were dissected out. The remaining carcass was cut in small pieces and was placed into 2ml/embryo of trypsin in 15ml tube (FALCON). The tube was left at 4°C for 2-18 hours and then at 37°C for 10 minutes. The embryonic tissue was pipetted up and down every 5 minutes and equal volume of EF medium was added. The big pieces of embryonic tissue were left to settle for 2 minutes, the supernatant was removed and plated out in one 162cm^2 tissue culture flask in 50ml of EF medium. The flask was incubated at 37°C and 5% CO_2. The medium was changed the next day and the fibroblasts were left to grow for 2 further days until they reached extreme confluency. The fibroblasts from one flask were split 1/5 and passed into 5 new 162cm^2 flasks. When they reached confluency after 3 days, the fibroblasts from each flask were split 1/5 and were passed into five

$162cm^2$ flasks again. They were left to grow for further 3 days and then the fibroblasts from all 25 flasks were collected, γ-irradiated (3500 rads) and were frozen in EF medium containing 12% DMSO at $4x10^6$ cells/vial or $20x10^6$ cells/vial. The total yield of fibroblasts is $400\text{-}600x10^6$ irradiated EFs/embryo. The irradiated EFs were used to form a monolayer on the culture dish in which the ES cells were grown. One vial of $4x10^6$ cells is enough for 1x10cm plate or 2x6cm plates.

EF Medium:

450ml Dulbecco's Modified Eagle's Medium (DMEM) with 0.11g/l NaPyruvate, pyridoxine; 50 ml Foetal calf serum (FCS); 5ml Glutamine 100x stock at 200mM; 2.5ml Penicillin/ Streptomycin 200x stock at 10^4 units penicillin and 10mg/ml streptomycin.

Transfection and selection of ES cells

The ES cells were grown to 50-100% confluency. The DNA used for the transfection was a $CsCl_2$ banded plasmid, linearised, ethanol precipitated and redisolved in sterile TE at concentration 2mg/ml. 10cm and 6cm culture dishes with $4x10^6$ and $2x10^6$ embryonic fibroblasts respectively were prepared in advance to receive the electroporated cells. The ES cells were trypsinised, spun out through medium and resuspended in Hepes buffered saline (HBS) containing 0.0007% 2-mercaptoethanol (2-ME) at a concentration of 10^7cells/ml. The linearised targeting construct was added to the stock of cells at a final concentration of 45μg/ml or 50μg/ml and 0.8ml of this mix was used for each electroporation. The electroporations were done with a BIORAD Gene Pulser at 400Volts and 25μF using 0.4cm electrode gap micro-

cuvettes (BIORAD). The cells were transferred then to a 15 ml Falcon tube containing 3.4 ml medium and plated out onto 4x10cm plates (0.9ml/plate). Selection was initiated 24hrs later at a concentration of 250μg/ml or 300μg/ml G418.

The selective medium was changed every 3 days and 576 resistant colonies were picked after 10 days onto 96-well plates, in 20μl PBS using P-200 Gilson pipette and tip. 6 days later the clones were transferred to 24-well plates and expanded for another 5 days. At confluency, half of each well was frozen in ES medium containing 12% DMSO and the other half was used to make DNA.

Hepes buffered saline (HBS):

137mM NaCl; 5mM KCl; 0.7mM Na_2HP0_4; 6mM dextrose; 21mM Hepes, pH7.1.

G418 stock:

50mg/ml in PBS stored up to a month at 4°C.

Gancyclovir stock:

1mg/ml in PBS stored frozen.

Karyotyping of ES cells

The cells were grown to 50-75% confluency, the medium was changed and 2 hours later Demecolcine (Sigma) was added at a final concentration of 20ng/ml to arrest the cells in metaphase. Two hours later the cells were trypsinised for 3 minutes at 37°C, spun out through ES medium for 5 minutes at 1200rpm and washed one more time in PBS. They were then resuspended in 5ml KCl 0.56% (Hypotonic solution) and let

stand at room temperature for 6 minutes following centrifugation at 750rpm for 5 minutes. The pellet was resuspended carefully in 5ml fixative by adding it drop wise, flicking thoroughly, left at room temperature for 5 minutes, and spun down for a further 5 minutes. The fixing procedure was done another 3 times and finally the cells were resuspended in 1ml fixative and dropped from 30cm height on a microscope slide that was washed before in 5% Acetic Acid in ethanol. The metaphase spreads were stained with 10% Giemsa's (DBH) for 20 minutes and chromosomes were count on a high power microscope.

Fixative:

Methanol: Acetic Acid 3:1

Preparation of primary cells for culture

Tissues were placed directly into cold air-buffered (AB) IMDM upon removal. Single cell suspensions were obtained either using a 0.2mm mesh (Becton-Dickinson) or by gentle teasing with forceps. The cells were centrifuged at 1200 rpm for 7 minutes at room temperature and resuspended in appropriate medium.

Complete RPMI medium:

450ml RPMI 1640 (GibcoBRL); 10% FCS; 200μM L-Glutamine; 2.5ml 100x stock Streptomycin (10mg/ml) / Penicillin (104 units); 50μM β-mercaptoethanol.

Purification of Peripheral T cells

A maximum of $3x10^6$ cells were loaded on to a T cell enrichment column (R&D Systems) following the protocol supplied by the manufacturer. The eluate containing the purified cells was centrifuged at 1200 rpm for 7 minutes at room temperature and resuspended in 2ml complete RPMI. Purity of the samples was determined by flow cytometry using anti CD4 and anti CD8 antibodies. Purity obtained was always >90%.

T cell proliferation assay

96well plates were pre-coated with serially diluted anti-CD3 (2c11) in PBS, with a starting concentration of 25µg/ml. The plates were incubated overnight at 4°C. Prior to use, the plates were washed three times with PBS. Purified primary cells were incubated with CFSE (Sigma) for 10 minutes at room temperature. The prepared cells were then plated in complete RPMI at $2x10^5$ cells per well, in a total volume of 200µl, and incubated at 37°C for up to 72 hours. The samples were then analysed by FACS analysis. Proliferation was assessed by measuring the proportion of cells in each peak. The calculations used to analyse the proliferation were as follows. Data of the frequency of CFSE labelled cells (F) that had undergone proliferation rounds (d) were used to calculate the adjusted frequencies (AdF), this was achieved by dividing F by 2^d. The percentage of cells triggered to divide (percent divided) was calculated by using the equation $(1-[AdF_0 / \Sigma AdF) \times 100$. The proliferation index was obtained from the calculation $\Sigma(AdF \times d) / \Sigma F_{(\text{for } d>0)}$.

Animal work

Transgenesis

Before microinjection of a DNA construct, the prokaryotic sequences were removed by restriction enzyme digestion followed by a purification of the construct using agarose gel electrophoresis as described before. The purified DNA was further cleaned on a Schleicher and Schuell Elutip column according to manufacturer's instructions.

Fertilized oocytes from (CBA/Ca X C57BL/10) F1 mice were isolated, and pronuclei were injected with a solution of the isolated DNA fragment at a concentration of 2-4 mg/ml in TE buffer. Eggs surviving microinjection were transferred into oviducts of recipient pseudo pregnant females. Microinjection of DNA constructs and transfers were carried out by Mauro Tolaini.

Injection of ES cells into blastocysts and generation of mutant mice

The cells were grown to 70-80% confluency, then trypsinised for 3 minutes at 37°C, spun out through 5ml ES medium at 1200rpm for 5 minutes and resuspended in 5ml Injection medium. 3.5-day blastocysts were collected from C57Bl/6 females and kept in blastocyst medium. 15-20 ES cells were injected into each blastocyst and 10-12 blastocysts were transferred into the uterus of each foster mother. Chimeric males were bred to C57Bl/6 females and germline transmission was determined by agouti

coat colour. Injection of blastocysts and transferring into the uterus were carried out by Mauro Tolaini and Dimitris Kioussis.

Injection medium:

Dulbecco's modification of Eagle's medium with 20mM Hepes, without glutamine and bicarbonate (ICN Biomedicals); 10% Foetal calf serum;1.5% Penicillin/streptomycin/ glutamine.

Blastocyst medium:

Dulbecco's modification of Eagle's medium; 10% Foetal calf serum; 1.5% Penicillin/ streptomycin/ glutamine.

Fluorescent Microscopy

Whole Mouse Microscopy

Photographs were taken using a Nikon coolpix 5000 digital camera mounted on a Leica MZ FLIII fluorescence stereomicroscope fitted with a GFP fluorescence filter.

Cell Microscopy

To perform the EYFP fluorescence experiment, HEK293 cells were grown on cover slips in DMEM medium. The cells were transfected with 1 μg of desired plasmids using Lipofectin (LifeTechnologies, Inc) according to manufacturer's instruction. 24 to 72 hours after transfection cells were fixed with 1.5 % Paraformaldehyde in PBS

for 15 min. EYFP images were analysed using a Deltavision microscope and the standard system operating software provided with the Deltavision microscope.

Flow cytometry

FACS Analysis

Single cells suspensions were prepared from thymus, lymph node and femoral bone marrow. To enrich for DN thymocytes, $CD4^+$ and $CD8^+$ thymocytes were depleted by staining single cell suspensions with anti-CD4 (RL172.4) and anti-CD8 (AD4(15)) followed by treatment with guinea pig complement (Life Technologies) and rabbit complement (Sigma). The resulting viable cells were recovered on a ficol gradient. For flow cytometry 10^6 cells were stained in 100µl ice-cold PBS supplemented with 0.5% BSA and 0.02% sodium azide. Cells were incubated for 30 minutes on ice with Fc blocker to minimize non-specific binding prior to a 30 minutes incubation with the specific antibodies. After the staining cell were washed twice with 100µl ice-cold PBS azide. Cells were again resuspended in 100µl PBS azide before data acquisition using a FACSCalibur and CELLQUEST software. All antibodies were used at appropriate saturating concentrations. The antibodies used were as follows: Fc blocker (anti-CD16/32), anti-Ter119-Biotin (Pharmingen), anti-Mac1-APC (Parmingen), anti-Gr1-Biotin (Pharmingen), anti-CD3-Biotin (Caltag), anti-CD19-APC (1D3), anti-IgD-Biotin (1.19), anti-CD4-PECY7 (Caltech), anti-CD8-APC (Caltech), anti-Thy1-PE (Pharmingen), anti-CD44-APC (Caltag), anti-CD25-Biotin (Pharmingen), anti-CD69-Biotin (Pharmingen), Streptavidin-PECY7 (Caltag).

Apoptosis assay (PI staining)

5 x 10^5 cells were used per sample. The cells were centrifuged at 1200 rpm for 5 minutes at room temperature, washed once with PBS and stained with propidium iodide (1mg/ml) in a hypotonic buffer overnight at 4°C in the dark. Analysis was carried out by flow cytometry as described by F. Nicoletti (Nicoletti et al., 1991).

Hypotonic buffer

0.85mM Na3Citrate; 0.1% Triton X-100

Electro Mobility Shift Assay

The shift assay was performed with a 36-mer double stranded oligonucleotide containing FT1 (5'-CTATTCAATCAATCCTTCATCGTTCATTCATTCAAC-3') or a double stranded oligonucleotide containing the binding site for OCT-1 (Promega). An aliquot of the single stranded oligonucleotides were annealed (94°C, 5 min; 58°C, 5 min) in 100 µl of TE buffer and then slowly cooled. The double-stranded DNA was 5' end-labelled with [γ^{32}P] ATP (ICN Pharmaceuticals, Inc.) and T4 polynucleotide kinase (Promega). After the labelling reaction, the unincorporated nucleotides were separated from the labelled DNA using MicroSpin™ S-200 columns (Pharmacia Biotech). DNA-protein binding reaction was performed as follows: 1µl of nuclear extracts (3µg/µl) was incubated with 9 µl Binding Buffer. In parallel, 0,5µl of probe (50,000 cpm of labelled, double stranded oligo) was incubated with 1µl of poly dIdC (200ng/µl) in a total volume of 10µl for 10 minutes. Subsequently, the two mixtures were added together and incubated for 20 minutes on ice. In a competition

experiment, a hundred times excess of unlabelled specific oligo nucleotide was incubated for 10 minutes before adding the specific radioactive probe. The labelled probe was separated from DNA-protein complexes by electrophoresis on 6 % non-denaturing polyacrylamide gels in 0.5 x Tris-borate/EDTA buffer pH 7.8 at RT. Autoradiography was carried out by exposure of the gel to intensifying screen for 6 to 14h at -70°C.

Binding Buffer:

400mM HEPES, pH 7.9; 200mM KCl; 10mM MgCl; 0.4mM EDTA; 1mM DTT; 30% glycerol

Proteins

Protein expression and purification

The construct sequences used in the present study are summarised in Table 1. ATX1 constructs were amplified from human ATX1 (EMBL/GenBank acc. no. P54253). HBP1 constructs were produced by PCR amplification of the M. musculus HBP1 sequence (acc. no. Q8R316). All constructs were produced as his-tagged GST fusion proteins. To obtain carboxy-terminal His and GST tagged recombinant proteins, the E. coli BL21(DE3) was transformed with DNA constructs. The single colony was picked from a freshly streaked selective plate, inoculated in a starter culture 100 ml LB supplemented with 100 μg/ml ampicillin and incubated overnight at 37°C. The pre-culture was transferred to a 6 litre of LB medium (100 μg/ml ampicillin, 5 mg/ml tetracycline and 30 μg/ml chloramphenicol) and further incubated until the A_{600} was

equal to 0.5. The protein expression was induced by adding isopropyl-β-D-thiogalactoside (IPTG), to a final concentration of 1 mM and incubated for 4-6 h. After that time, cells were harvested by centrifugation at 4°C for 15 min at 5000 rpm. The bacterial pellet from a 6 L culture was resuspended in BugBuster reagent by pipetting or gentle vortexing, using 5 ml reagent per gram of wet cell paste. 1 µl (25 units) Benzonase Nuclease was added per 1 ml BugBuster reagent used for resuspension. The cell suspension was incubated on a rotating mixer at a slow setting for 10–20 min at room temperature. Insoluble cell debris was removed by centrifugation at 13,000xg for 20 min at 4°C. The supernatant was applied to a Ni-NTA agarose column (QIAGEN). Cleavage of the his-tagged GST was achieved overnight at room temperature using GIBCOBRL His-tagged TEV recombinant protease. Separation of GST and TEV from the proteins was performed using a Ni-NTA agarose column. The cleavage product was recovered by elution.

Preparation of nuclear extracts

Nuclear extracts from mouse thymus were prepared as described (Dignam et al., 1983). Briefly, the tissue was disrupted into smaller pieces and about 10^9 cells were spun down at 3,000 rpm, 10min, Beckman J6 at 0°C. The pellet was resuspended in ice-cold homogenisation buffer in 5 times the volume of the pellet and homogenised with 10 strokes using a dounce tissue grinder (Wheaton). 3.5 ml of the homogenate were laid carefully on the top of the 1.5 ml pre-made cushions of homogenisation buffer in 5 ml tubes of the SW55 rotor. After centrifugation (100,000xg/30,000rpm, 0°C, 1 hour) the supernatant was removed and the nuclear pellet was resuspended in 3.5 ml per tube nuclear lysis buffer. The nuclei were lysed adding 0.04 volumes of

0.5% SDS and 0.1 volume of 4M $(NH_4)_2SO_4$. The lysate was spun in the SW55 rotor (150,000xg/40,000rpm, 0°C, 1hr). The supernatant was transferred into a fresh tube and proteins were precipitated by adding solid $(NH_4)_2SO_4$ to a concentration of 0.33mg/ml. The protein pellet was collected by centrifugation in the SW55 rotor (100,000xg/30,000rpm, 0°C, 20 minutes). The pellet was resuspended in 1 ml of nuclear dialysis buffer and dialysed against >200 volumes of Dignam nuclear dialysis buffer for 2hr at 4°C. The concentration of protein extract was measured by a spectrophotometer at 590nm (against titrated BSA). Subsequently the extract was aliquoted into small samples, frozen and stored at -70°C.

Homogenisation buffer:

10mM Hepes, pH 7.6; 15mM KCl; 1mM EDTA; 0.1mM EDTA; 2.2M Sucrose; 10mM NaF

Nuclear Lysis Buffer:

10mM Hepes, pH 7.6; 100mM KCl; 3mM $MgCl_2$; 0.1mM EDTA; 10% Glycerol (v/v); 10mM NaF

Dignam Dialysis buffer

20mM Hepes, pH7.6; 100mM KCl; 0.2 mM EDTA; 20% Glycerol (v/v); 0.5mM NaF

Results

Structural analysis of AXH module

Properties of the AXH module

As described in the introduction of this thesis a small motif of roughly 130 amino acids, named AXH module, was identified in the apparently unrelated factors HBP1 and ATX1. Protein motifs are conserved sequences within several protein families. The conservation of the motif suggests that it participates in a common structural element and/or a common function that is critical for the proteins in which it was found. Thus, the homology found between ATX1 and HBP1 could provide insights into the function of both proteins. We, therefore, set up a collaboration with the department of Molecular Structure at the National Institute for Medical Research to characterize the structural properties and stability of this domain. My part of this work consisted of the cloning and expression of the protein domains.

Defining the domain boundaries

In order to determine the minimal length required to obtain a folded, soluble and highly expressed AXH from both ATX1 and HBP1, constructs for each of the two proteins were initially designed (ATX1 (568-689) and HBP1 (208-345)) which

encompass the region of maximal homology (Table 1). Both proteins were expressed in good yields with final yields of around 20mg/litre of culture. The secondary and tertiary structures were checked by circular dichroism (CD) and nuclear magnetic resonance (NMR) respectively (The CD and NMR experiments were done by Cesira de Chiara and will not be further discussed in this thesis). To check whether the AXH module could still fold properly if cut further, constructs ATX1 (574-689) and HBP1 (208-335) were designed, omitting an N-terminal and a C-terminal proline rich stretch respectively (Table 1). The rationale for this choice was that both stretches are not conserved and their sequences do not suggest secondary structure elements. ATX1 (574-689) was isolated from the soluble fraction but the yields of the recombinant protein were extremely low yields (less than 1mg/litre of culture). Both the CD and 1D NMR spectra of the protein showed that the fragment retains some 3D fold but adopts a conformation different from that of the longer version (The CD and NMR experiments were done by Cesira de Chiara). In contrast, most of the expressed HBP1 (208-335) formed inclusion bodies despite all attempts to shift the equilibrium towards the soluble form. To investigate a possible role of the flanking residues in domain stability, one more construct was designed for each of the two proteins by extending the boundaries both N- and C-terminally. In the absence of other indications, we selected the new boundaries to include predicted secondary structure elements. The NMR spectra of HBP1 (200-343) and ATX1 (556-700) are those of folded species and reminiscent of those of HBP1 (208-345) and ATX1 (568-689). However, the spectrum of ATX1 (556-700) changes with time, leading to a major line broadening indicating that the protein tends to aggregate. These results show that the stability of AXH domain is mostly independent from flanking regions but depends on

the domain integrity. Importantly, mutants excluding residues within the domain boundaries defined here lead to unfolded proteins.

The AXH module of ATX1 but not of HBP1 forms dimers

The ATX1 and HBP1 constructs were then probed for their association state by analytical ultracentrifugation (The analytical ultracentrifugation experiments were done by Cesira de Chiara and only the final result will be described). The apparent molecular weight of 18.0 kDa derived for HBP1 (208-345) AXH domain approximates within experimental error the monomer value (16,410). Sedimentation equilibrium curves of ATX1 (568-689) AXH domain can be fit either as a dimer (at 200 mM NaCl) or as a mixture of a prevalent dimeric species in equilibrium with further aggregates (at 50 mM NaCl). A higher tendency to aggregate was observed for the extended ATX1 (556- 700) AXH domain which, depending on salt conditions, is either a tetramer or a multimer in solution, confirming that extension of AXH of ATX1 promotes further aggregation. In contrast, HBP1 (200-343) AXH domain remains monomeric. These results show that the AXH domain in isolation is in a different aggregative state in the two protein families: while the AXH of HBP1 exists as a monomer, the AXH of ATX1 exists in solution as a dimmer or aggregates.

Table 1 Summary of the constructs used in the study

Constructs	Boundaries	Behaviour
HBP1(208-335)	PSTVWHCFLK.....VCL	Inclusion bodies
HBP1(208-345)	PSTVWHCFLK.....VCLPPGHPDAINF	Folded, monomer
HBP1(200-343)	DLGWCNSWPSTVWHCFLK.....VCLPPGHPDAI	Folded, monomer
ATX1(556-700)	PVVQSBASPAAAPPTLPPYFMK.....VCISLTLKNLKNGSVKKGQ	Aggregating
ATX1(568-689)	PPTLPPYFMK.....VCISLTLKNLK	Folded, dimer
ATX1(574-689)	YFMK.....VCISLTLKNLK	Low yield

HBP1 intracellular localization.

To gain insight into those sequences in the HBP1 that mediate nuclear import, we performed sub-cellular localization studies on wild type and mutant proteins in living cultured cells. First, we determined the sub-cellular localization of wild type HBP1. For that purpose, the complete HBP1 coding sequence was fused to the C-terminus of EYFP using the pEYFP-C1 vector (BD bioscience) (figure 7A). This approach has the advantage that localization of fusion proteins can be monitored directly by fluorescence microscopy. These constructs were transfected transiently into human embryonic kidney cell line 293 and analyzed for EYFP expression 24 hours post-transfection. As published previously (Stauber et al., 1995), EYFP alone localized in both the cell nucleus and the cytoplasm (figure 7B). This is because proteins smaller than 60 kDa are able to enter the nucleus passively by diffusion through the nuclear pores. In sharp contrast, however, expression of HBP1–EYFP fusion protein resulted in exclusively nuclear accumulation, highlighted by a significant concentration in nucleoplasmic speckles (figure 7B). Given its size (around 80 kDa), this fusion protein is not able to enter the nucleus passively by diffusion through the nuclear pores. This might indicate that active processes may be involved in the import of HBP1 into the nucleus.

In order to define the regions of the HBP1 protein that are essential for nuclear transport of HBP1, fusion proteins consisting of EYFP and various HBP1 domains were constructed. It is common for nuclear proteins that their nuclear localisation

domain co-localises with their DNA binding domain. Therefore, we constructed two different fusion proteins. The first construct consists of the HMG box, which is the DNA binding domain of HBP1, fused to EYFP from now on referred to as EYFP-HMG. The second construct consists of the entire N terminal region of HBP1 up to the HMG box fused to EYFP from now on referred to as EYFP-ΔHMG (figure 7A). Transfection of the constructs into human embryonic kidney cell line 293 cells led to expression of the respective recombinant proteins. Fluorescence microscopy of the various transfected cell-lines revealed that the EYFP-HMG fusion protein

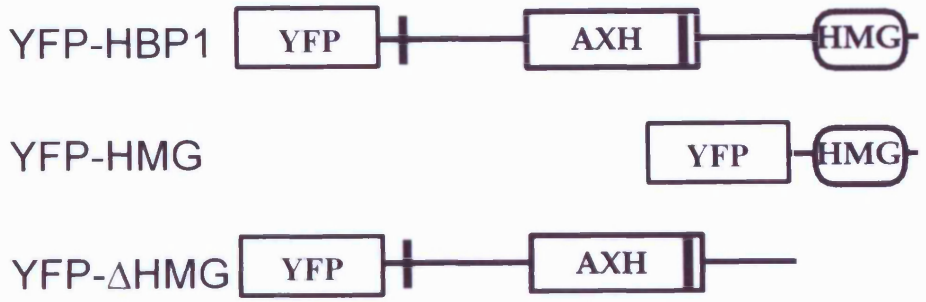
was located exclusively in the nucleus (figure 7B), while the construct in which the HMG box was deleted (EYFP-ΔHMG), was found throughout the cell giving the same staining pattern as EYFP alone (figure 7B). The size of this fusion protein was around 60 kDa. This allows the fusion protein to enter the nucleus passively by diffusion through the nuclear pores. This might explain why the protein is not found exclusively in the cytoplasm.

It should also be noted that a nucleoplasmic speckle pattern was observed when the EYFP-HMG box fusion protein was transfected into the cells. However, this nucleoplasmic speckle pattern showed smaller speckles compared to the wild-type HBP1 (figure 7B).

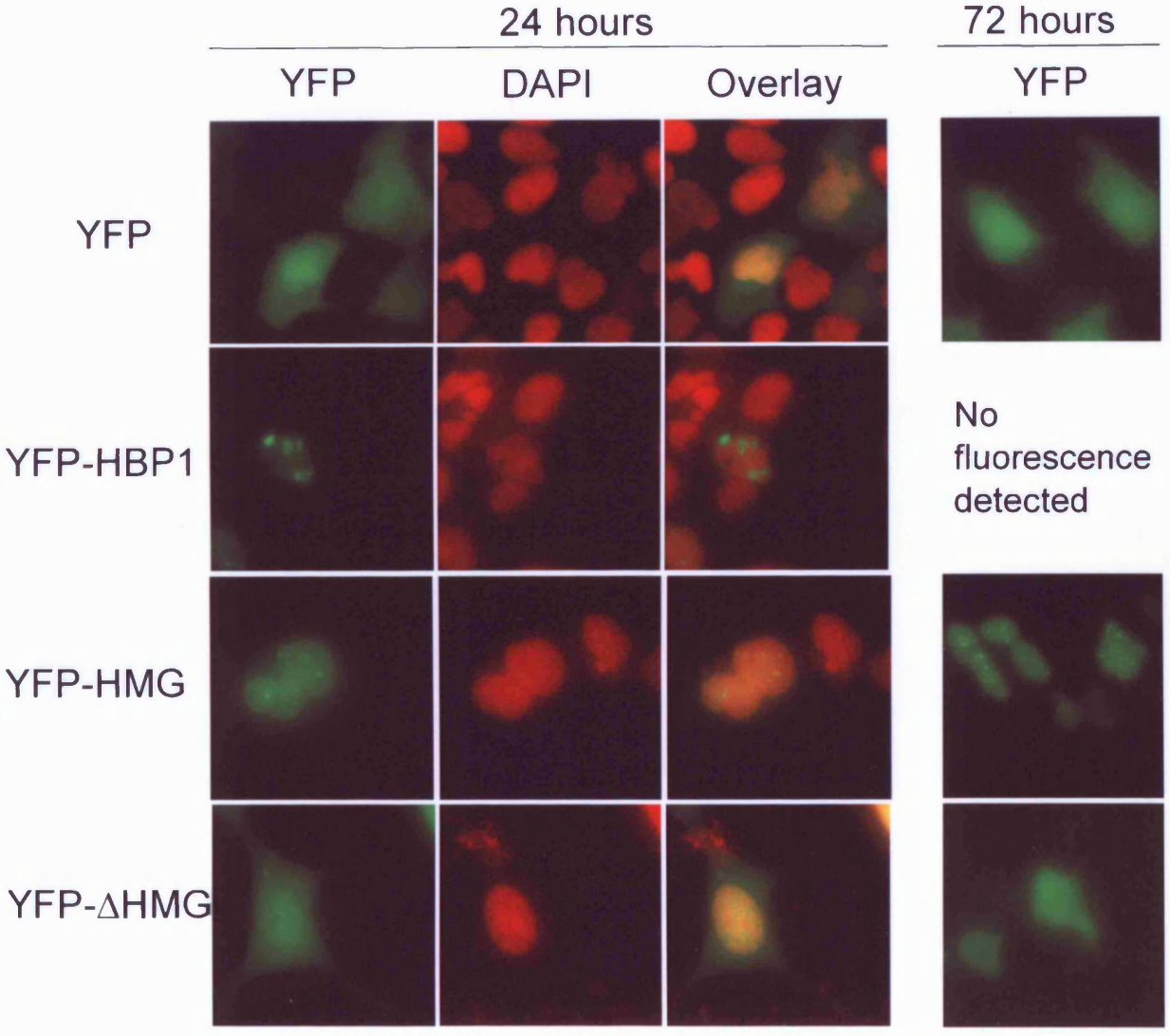
Figure 7: The subcellular localization of full length HBP1 and truncated HBP1 in HEK-293 cells

A) Schematic representation of the EYFP-HBP1, EYFP-ΔHMG and EYFP-HMG. B) The fluorescence image of HEK293 cells transfected with each EYFP fusion protein. EYFP expression can be visualized directly by its intrinsic green fluorescence. Counterstaining is with the nuclear dye DAPI (red fluorescence).

A: Constructs



B: YFP expression



When the cells were analysed at 72 hours post transfection the distribution of EYFP-HMG and EYFP-ΔHMG was the same as at 24 hrs post transfection (figure 7B). For the full-length EYFP-HBP1 fusion protein at 72 hrs post transfection, no fluorescent living cells were detected. Time-lapse videos showed that cells expression the full-length EYFP-HBP1 fusion protein changed there morphology (cells became round) and detached from the cover slip (data not shown). Rounding up of the cell and detachment from the matrix are signs of apoptosis (Nakayama, 1998). This might indicate that over-expression of the EYFP-HBP1 protein induces cell death.

Taken together these results indicate and confirm that the HMG box sequences contain a nuclear localization signal that leads to the import of the transfected protein into the nucleus.

Generation and analysis of Cre transgenic mice.

Establishing Cre transgenic mouse lines

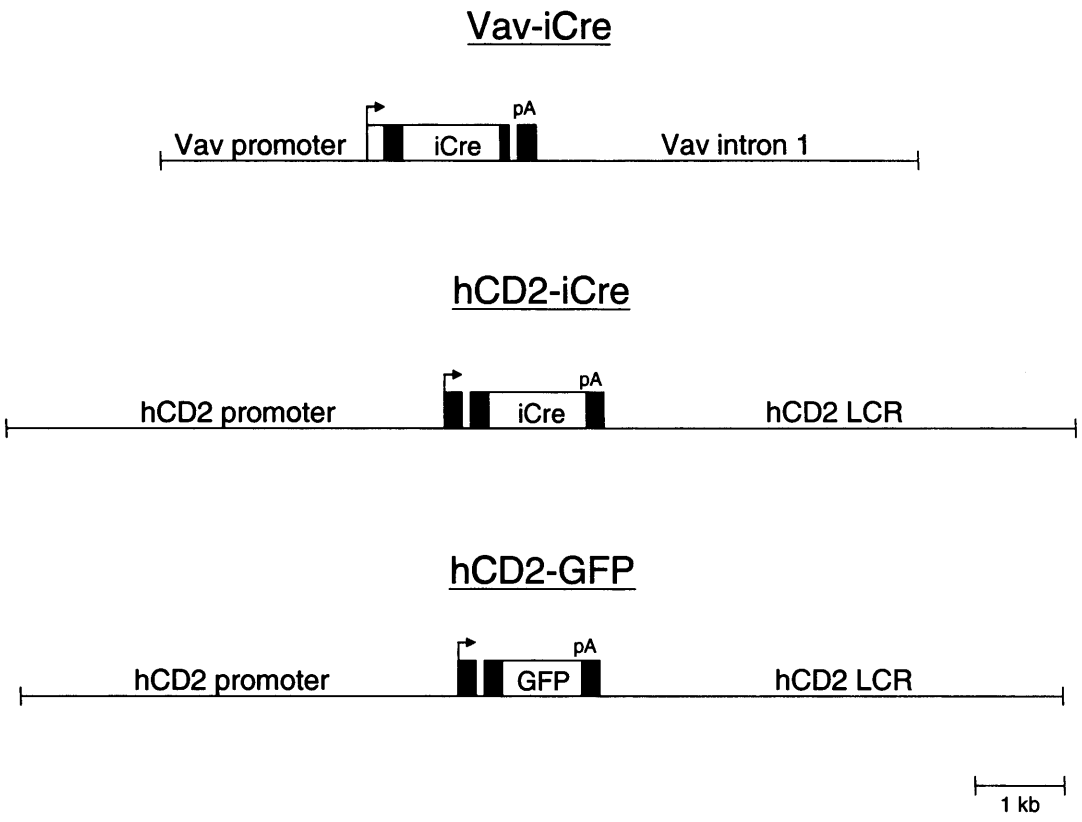
In order to generate a system for conditional gene targeting within the myeloid and lymphoid lineages two different regulatory elements were chosen to direct expression of iCre in transgenic mice. The improved Cre (iCre) gene was chosen instead of the "normal" Cre because iCre contains a reduced CpG content (Shimshek et al., 2002). High CpG content of prokaryotic coding sequence has been associated with epigenetic silencing in mammals. This part of this thesis was done in collaboration with Adam Williams.

To restrict the iCre expression to the haematopoietic compartment the vav1 expression vector was used. The vav1 expression vector (HS21/45 *vav*-hCD4) has been shown previously to drive expression of a hCD4 transgene throughout the entire haematopoietic compartment (Ogilvy et al., 1999). Vav1 is a signal transducer protein expressed exclusively in the haematopoietic system, where it plays a pivotal role in growth factor-induced differentiation and proliferation. Vav1 couples tyrosine kinase signals with the activation of the Rho/Rac GTPases, leading to cell differentiation and/or proliferation (Tybulewicz et al., 2003). Thus, to restrict conditional targeting to the haematopoietic system, the hCD4 gene was removed from the vav1 expression vector and replaced with the iCre cDNA (done by Adam Williams) (figure 8). To restrict expression further to lymphocytes, the iCre cDNA was inserted into, the earlier described, hCD2 VA vector, which has previously been shown to drive high levels of transgene expression in T cells and occasionally in B cells, but not in other tissues (Lang et al., 1988). As a control, the GFP open reading frame was inserted directly into the hCD2 VA cassette (figure 8).

After linearization and removal of plasmid sequences each of the Vav-iCre, hCD2-iCre and hCD2-GFP constructs were injected into the pronuclei of fertilised oocytes (injections were done by Mauro Tolaini). The presence of the transgene in founders was determined by Southern blot analysis. After germ line transmission a minimum of four independent mouse lines were established for each construct. Of each of the Vav-iCre, hCD2-iCre and hCD2-GFP lines, one line of each which shows wildtype number of haematopoietic cells (table 2 on page 122) and good reporter gene expression (data not shown) was chosen for further analysis.

Figure 8 Cre expression constructs

The iCre cDNA (Shimshek et al., 2002) was inserted into the HS21/45 vav-hCD4 expression cassette (Vav-iCre) (Ogilvy et al., 1999). The untranslated portion of the Vav1 exon 1 is represented by a light grey box. The SV40 splice sites and poly-adenylation signal (pA) are represented by dark grey boxes. The iCre and GFP cDNAs were inserted into the hCD2 VA cassette (Zhumabekov et al., 1995) (hCD2-iCre and hCD2-GFP, respectively). Exon 1, exon 2 and the hCD2 poly-adenylation signal (pA) are represented by black boxes.


Vav-iCre
pA
Vav promoter
iCre
Vav intron 1
hCD2-iCre
pA
hCD2 promoter
iCre
hCD2 LCR
hCD2-GFP
pA
hCD2 promoter
GFP
hCD2 LCR
1 kb

To determine the efficiency of iCre expression, Vav-iCre and hCD2-iCre transgenic mice were crossed to the homozygous R26R-EYFP reporter mouse (Srinivas et al., 2001). The R26R-EYFP reporter line contains an EYFP transgene inserted into the ROSA26 locus by homologous recombination. EYFP is only expressed when a floxed transcriptional stop sequence that prevents transcription from the ROSA26 locus is excised by Cre mediated recombination (figure 9). Resulting EYFP expression can be detected by either fluorescence activated cell sorting (FACS) or by fluorescence microscopy.

Tissue restricted expression of iCre.

In order to investigate iCre expression patterns in each of the selected lines, mice transgenic for both iCre and EYFP were dissected under a fluorescence microscope fitted with a GFP filter. Mice either transgenic only for EYFP or for hCD2-GFP were also dissected as negative and positive controls, respectively (figure 10).

In Vav-iCre/EYFP and hCD2-iCre/EYFP transgenic mice, superficial lymph nodes were obvious and appeared bright and uniform in colour and fluorescent intensity (figure 10). Other lymphoid tissues such as the thymus, spleen, Peyer's patches and intestinal crypto patches were also positive for EYFP expression (figure 10). Further dissection revealed that the mesenteric (figure 10) and inguinal lymph nodes (data not shown) were positive for EYFP expression. Analysis of hCD2-GFP transgenic mice showed an almost identical distribution of fluorescence, as observed in Vav-iCre/EYFP and hCD2-iCre/EYFP transgenic mice. In contrast, however, to the

Figure 9 Fate mapping cells that have undergone recombination

A tissue-specific promoter (In our case hCD2 or VAV) controls the expression of iCre transgene. These mice were crossed to a reporter line, which has EYFP cDNA inserted into the ubiquitously expressed ROSA26 locus, preceded by a loxP-flanked stop sequence. In cells that do not express iCre no EYFP is detected because Cre mediated recombination is required to remove the floxed Stop sequences that separate the promoter from the reporter coding sequences. Cre activity results in the removal of the Stop sequence and reporter activation. Even cells that no longer express Cre, will express the reporter.

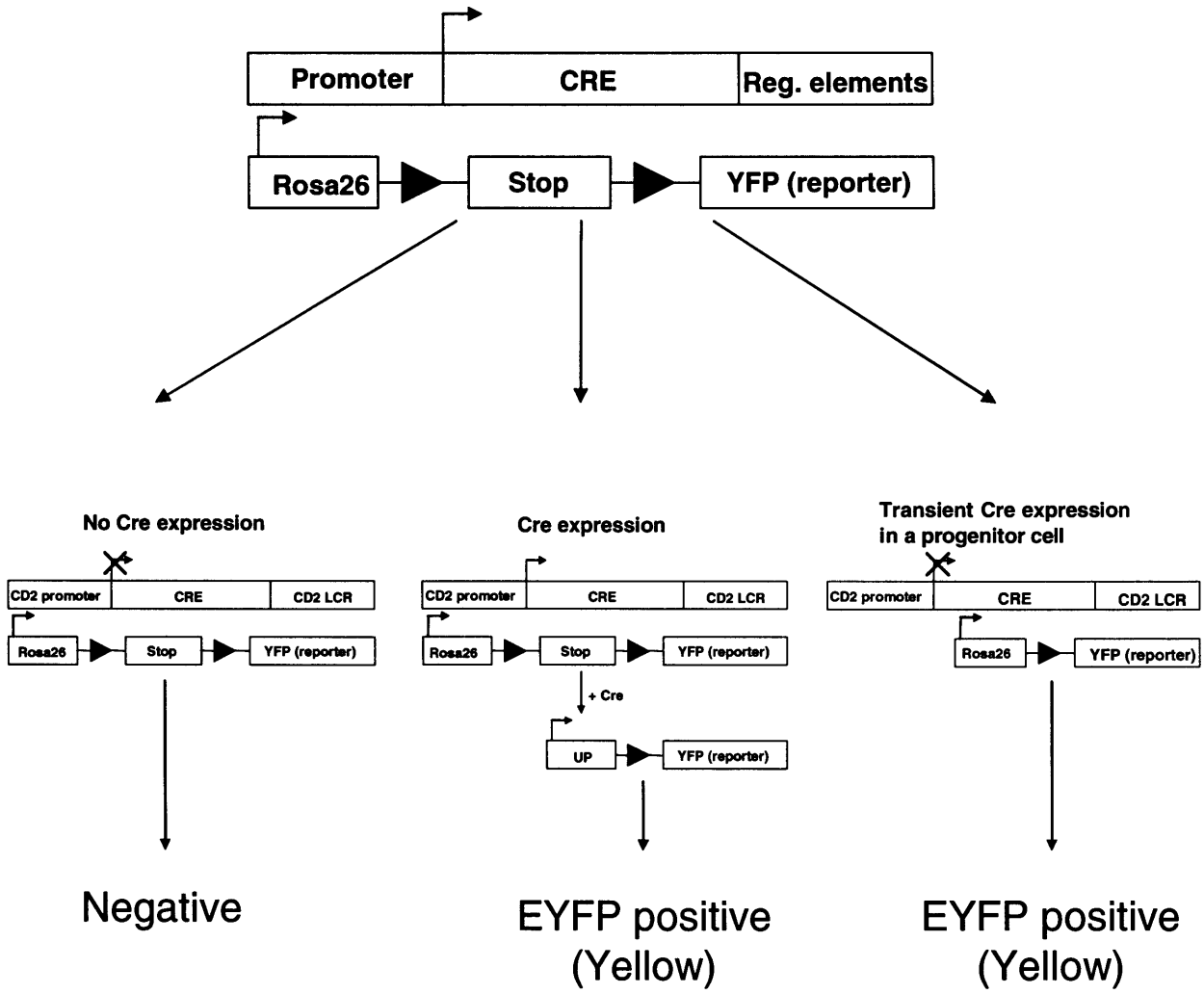
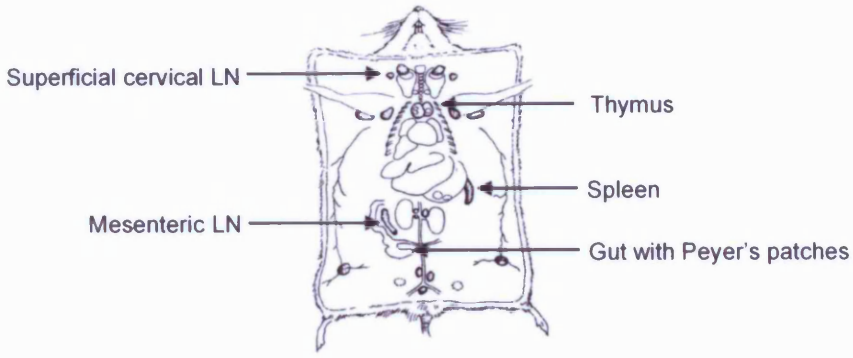
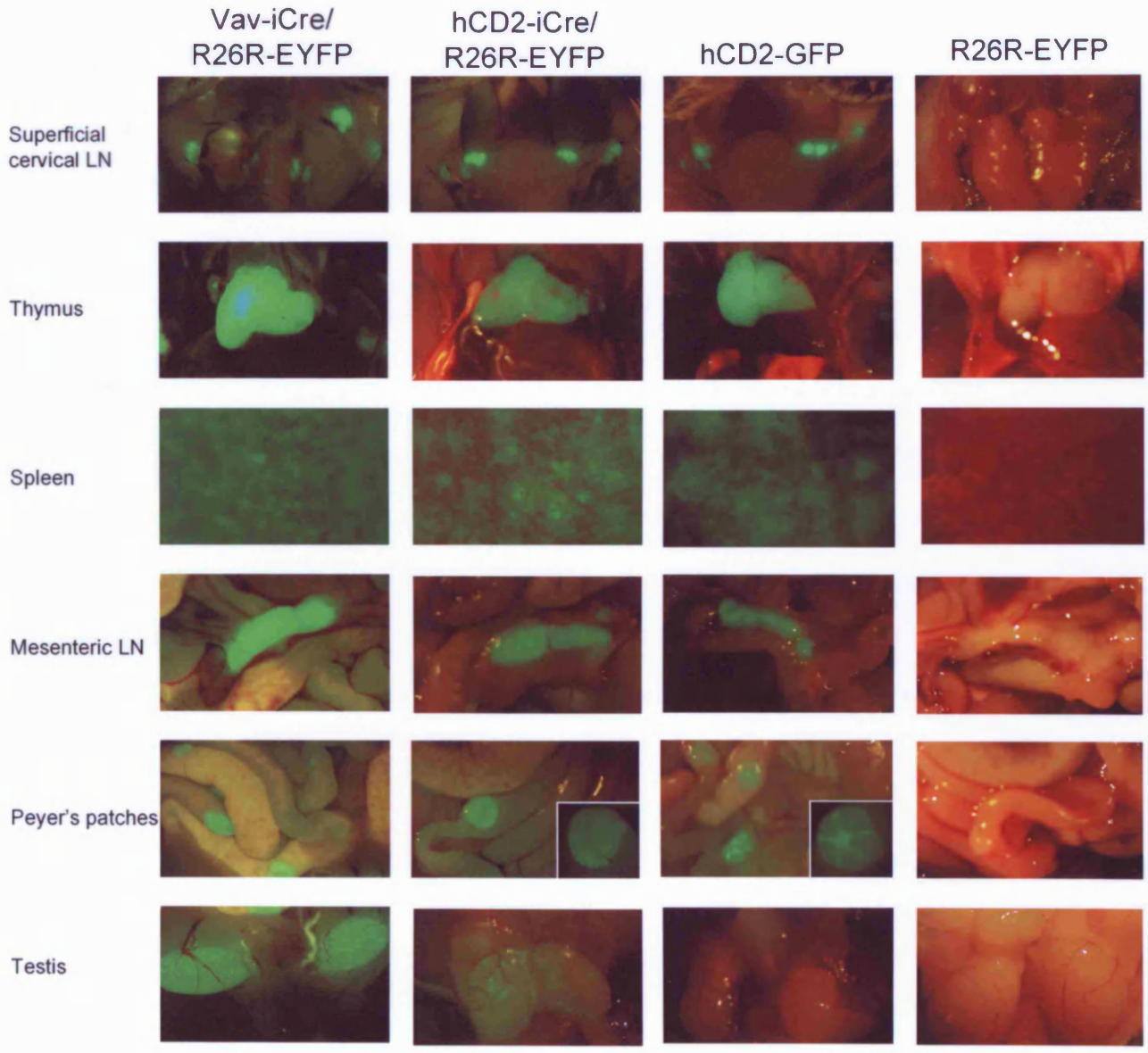

Promoter
CRE
Reg. elements
Rosa26
Stop
YFP (reporter)
No Cre expression
CD2 promoter
CRE
CD2 LCR
Rosa26
Stop
YFP (reporter)
Negative
Cre expression
CD2 promoter
CRE
CD2 LCR
Rosa26
Stop
YFP (reporter)
+ Cre
UP
YFP (reporter)
EYFP positive
(Yellow)
Transient Cre expression
in a progenitor cell
CD2 promoter
CRE
CD2 LCR
Rosa26
YFP (reporter)
EYFP positive
(Yellow)

Figure 10 Analysis of Cre expression patterns in tissues of transgenic mice by fluorescent microscopy

Vav-iCre R26R-EYFP double transgenic mice, hCD2-iCre R26R-EYFP double transgenic mice, hCD2-GFP mice were dissected under a fluorescent microscope fitted with a GFP filter. A R26R-EYFP single transgenic mouse is also shown as a negative control. Similar results were obtained in at least five independent experiments.


Superficial cervical LN
Thymus
Spleen
Mesenteric LN
Gut with Peyer's patches


Vav-iCre/ R26R-EYFP
hCD2-iCre/ R26R-EYFP
hCD2-GFP
R26R-EYFP
Superficial cervical LN
Thymus
Spleen
Mesenteric LN
Peyer's patches
Testis

uniform expression as seen in Vav-iCre/EYFP and hCD2-iCre/EYFP transgenic mice, the Peyer's patches from hCD2-GFP mice exhibited areas of high and low fluorescence intensity (see below).

Surprisingly, expression of EYFP was noted in testis (figure 10) and ovaries (data not shown) of Vav-iCre/EYFP transgenic mice. Mosaic expression was also seen in testis of hCD2-iCre/EYFP transgenic males (figure 10), but not in ovaries of hCD2-iCre/EYFP females (data not shown). In both iCre expressing lines, all other tested tissues were negative for EYFP expression, indicating that, with the exception of testis and ovaries, expression of iCre was restricted to tissues of the haematopoietic and/or immune system. Interestingly, expression of GFP was not detected in testis or ovaries from hCD2-GFP mice (figure 10). No EYFP expression was detectable in any tissue analysed from EYFP single transgenic controls.

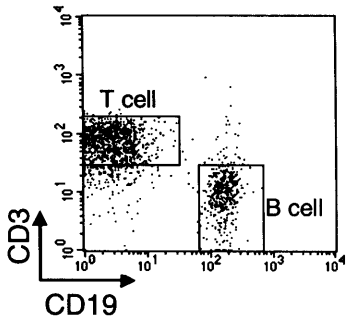
Cre expression in T cells

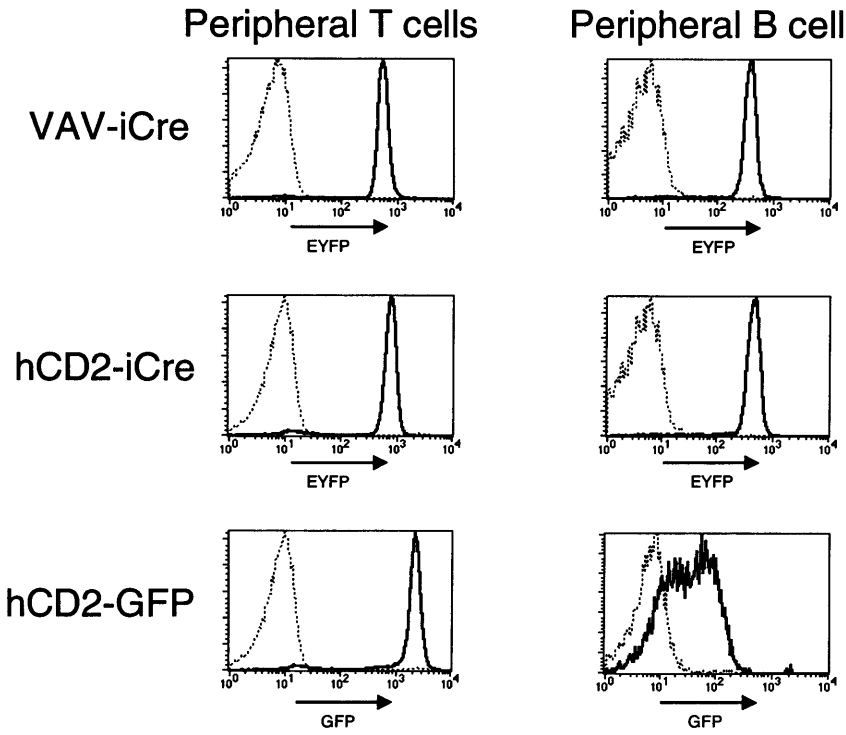
In order to identify the types of cells that have undergone Cre mediated recombination within the positive organs described above, cells isolated from lymph nodes were stained with anti-CD3 and anti-CD19 and analysed by flow cytometry (figure 11 and table 2 and 3). Analysis of both Vav-iCre/EYFP and hCD2-iCre/EYFP transgenic mice revealed that all peripheral $CD3^+$ T cells expressed EYFP. In addition, high levels of GFP expression were detected in all peripheral T cells from hCD2-GFP mice. Thymocytes were stained with anti-CD4 and anti-CD8 to distinguish the

Figure 11 Flow cytometric analysis of EYFP and GFP expression in lymph node derived lymphocytes

Single cell suspensions from lymph nodes were prepared from Vav-iCre R26R-EYFP double transgenic mice, hCD2-iCre R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice. These cells were stained with CD3 and CD19 antibodies to allow differentiation of the T cell (CD3+/CD19-) and B cell populations (CD3-/CD19+). Levels of EYFP or GFP expression on T cells and B cells are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the auto-fluorescence in non-transgenic controls. Each experiment was repeated at least three times with similar results.

Lymph node


T cell
B cell
CD3
CD19


Peripheral T cells
Peripheral B cells
VAV-iCre
hCD2-iCre
hCD2-GFP
EYFP
EYFP
EYFP
EYFP
GFP
GFP

$CD4^+/CD8^-$ or $CD4^-/CD8^+$ single positive (SP) thymocytes as well as $CD4^+/CD8^+$ double positive (DP) thymocytes (figure 12 and table 2 and 3). Detailed analysis revealed that Vav-iCre/EYFP and hCD2-iCre/EYFP transgenic mice exhibited high levels of EYFP expression on approximately 100% of both SP and DP thymocytes. This finding indicates that Cre was expressed in these cells or had been expressed at an earlier developmental time point. Identical analysis revealed a similar percentage of GFP expressing SP and DP thymocytes in the hCD2-GFP line. Importantly, no expression was seen on T cells derived from EYFP single transgenic controls (figure 11 and 12 dotted line).

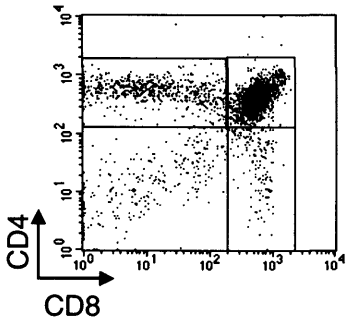
To determine precisely the time point at which Cre expression is initiated during thymocyte development in each of the lines, double negative (DN) thymocytes were stained with anti-CD44 and anti-CD25 in order to separate the four stages of DN thymocyte development (figure 13 and table 2 and 3). When thymocytes enter the thymus, they express high levels of CD44, but no CD25 (DN1 stage). As the thymocytes mature they up regulate CD25 to become $CD44^+/CD25^+$ cells (DN2 stage); subsequent down regulation of CD44 gives rise to $CD44^{low}/CD25^+$ thymocytes (DN3 stage). After successful T cell receptor β-chain rearrangement these cells down regulate CD25 to become $CD44^-/CD25^-$ (DN4 stage). Finally, these cells begin to express CD4 and CD8 to become DP thymocytes.

All four DN populations from the Vav-iCre/EYFP line showed high levels of EYFP expression in approximately 100% of cells, indicating that even the earliest T cell progenitors within the thymus had expressed Cre. Intriguingly, analysis of the hCD2-iCre/EYFP line revealed a slightly different pattern of EYFP expression. A population

Figure 12 Flow cytometric analysis of EYFP and GFP expression in thymocytes

Thymocytes from Vav-iCre R26R-EYFP double transgenic mice, hCD2-iCre R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were stained with CD4 and CD8 antibodies. Levels of EYFP or GFP expression on SP (CD4+/CD8- or CD4-/CD8+), DP (CD4+/CD8+) and DN populations (CD4-/CD8-) are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the auto-fluorescence in non transgenic controls. Each experiment was repeated at least three times with similar results.

Thymus



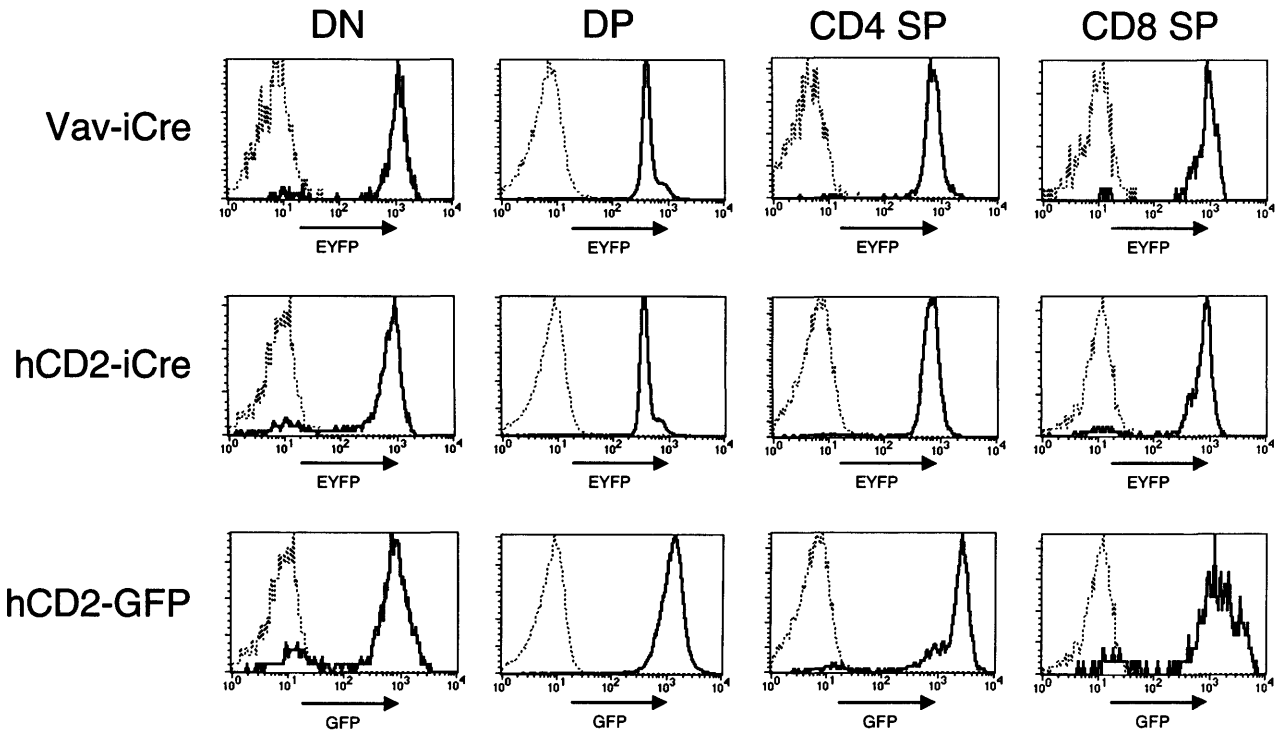
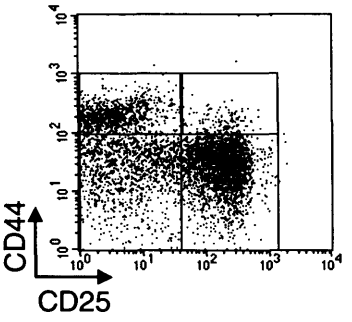


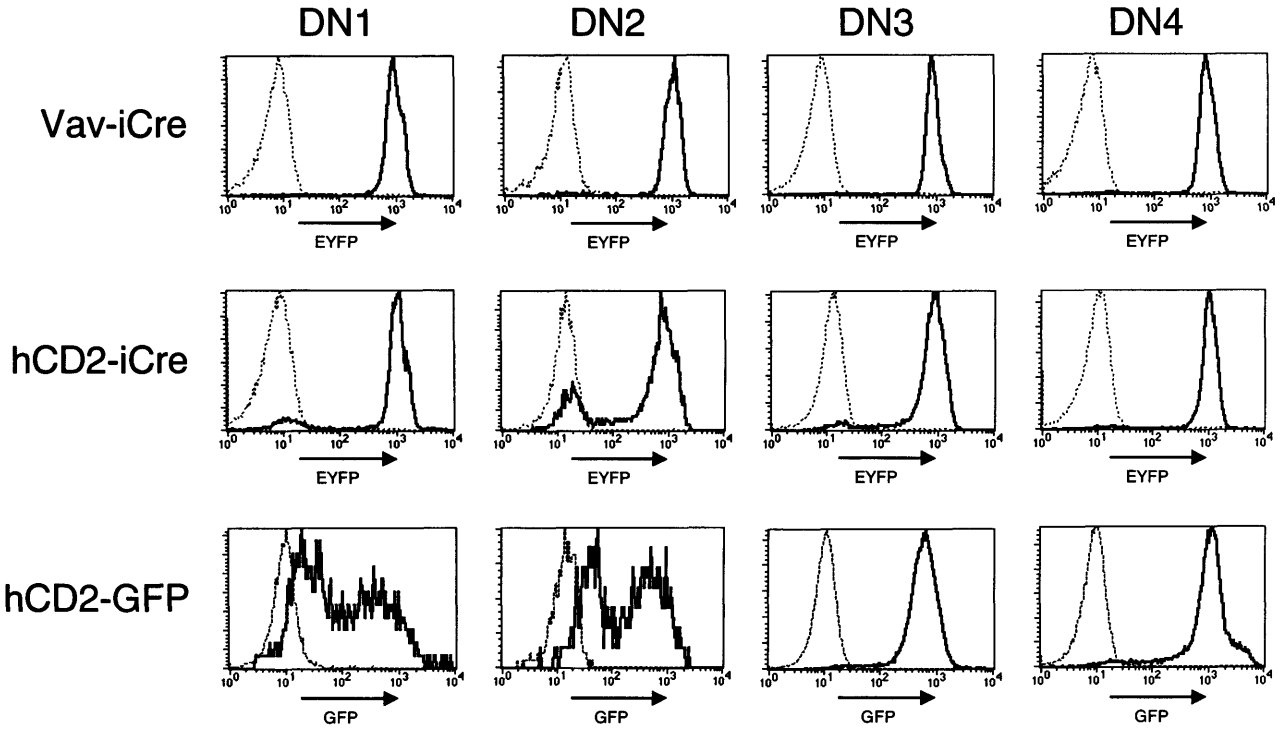
Figure 13 Flow cytometric analysis of EYFP and GFP expression during DN thymocyte development

Single cell suspensions from thymocytes from Vav-iCre R26R-EYFP double transgenic mice, hCD2-iCre R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were depleted from CD4 and CD8 expressing cells by complement treatment (see materials and methods). Remaining DP and SP thymocytes were excluded from analysis by staining with CD4, and CD8 antibodies. Resulting DN thymocytes were stained with anti-CD25 and anti-CD44 antibodies. Levels of EYFP or GFP expression at each DN stage are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the auto-fluorescence in non-transgenic controls. Each experiment was repeated at least three times with similar results.

Double negative thymocytes



DN1	DN2
DN4	DN3



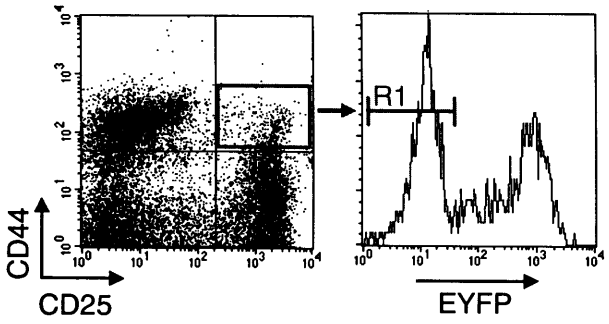
of EYFP negative cells were detected in DN1, DN2, and DN3 stages, which disappeared in the final DN4 stage. To determine if all the cells in DN2 were true immature T cells they were stained with a number of lineage markers (figure 14). The EYFP negative cells did not express the CD3/T-cell receptor complex as measured by anti CD3 as is true for DN thymocytes before they rearrange the TCR genes. They were negative for markers, which are characteristic for B cells (B220), Natural Killer cells (DX5), monocytes (Mac1), red blood cells (Ter119) and dendritic cells (CD11c). They were positive for the haematopoietic marker CD45 and expressed intermediate levels of the T cell marker Thy.1. Altogether, these data suggest that the negative cells in DN2 are true immature T cells that had not initiated Cre expression yet. Analysis of GFP expression in DN thymocytes from hCD2-GFP transgenic mice showed yet a different expression profile. A significant proportion of DN1 thymocytes did not express detectable amounts of GFP. As the thymocytes proceeded through to the DN2 stage the proportion of GFP expressing cells increased until the DN3 and DN4 stages, where 100% of the cells expressed GFP. These data suggest that the hCD2 transgenes start expressing sometime between the DN1 and DN3 stage and that expression is universal in all thymocytes by DN4 stage.

Cre expression in B cells

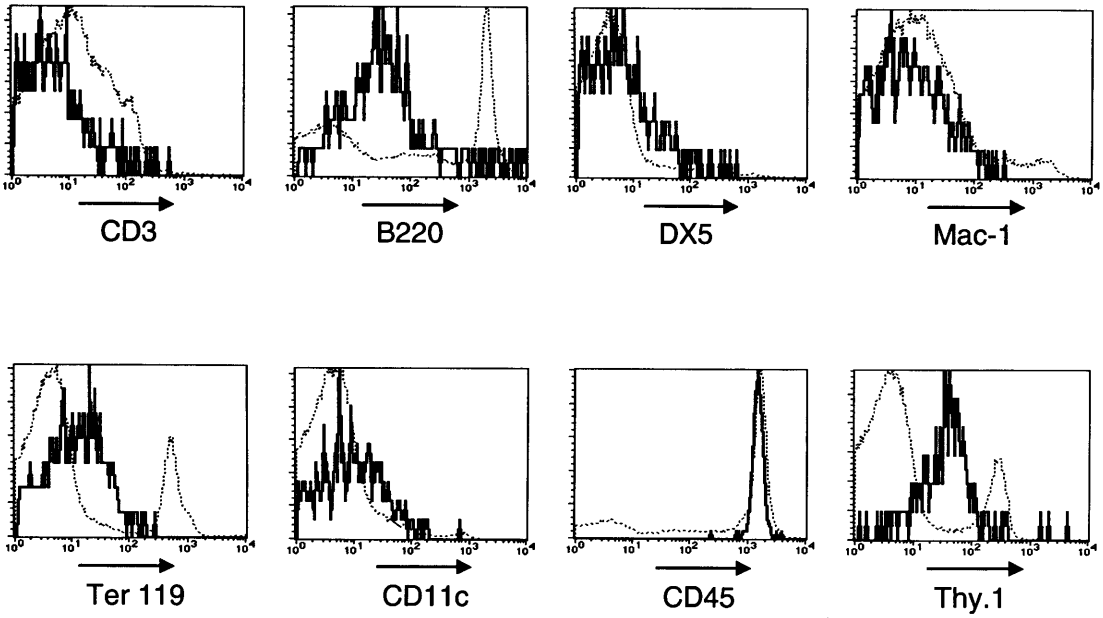
To analyse iCre expression in B lymphocytes, cells isolated from lymph nodes of Vav-iCre/EYFP and hCD2-iCre/EYFP transgenic mice were stained with anti-CD3 and anti-CD19 in order to distinguish the B cell and T cell populations (figure 11 and table 2 and 3). Analysis of EYFP expression revealed that almost all peripheral

Figure 14 Flow cytometric analysis of EYFP negative cells in DN2

Single cell suspension from thymocytes from hCD2-iCre R26R-EYFP double transgenic mice was depleted from CD4 and CD8 expressing cells by complement treatment (see materials and methods). The remaining DN thymocytes were stained with CD25 and CD44 antibodies and one of the following lineage markers anti-CD3, anti-B220, anti-DX5, anti-Mac-1, anti-Ter119 anti-CD11c, anti-CD45 or anti-Thy-1. The bold line in the histograms shows the expression profile for each marker on the EYFP negative cells in the DN2 compartment. The dotted line shows the staining profile of the used lineage markers on splenocytes.



Expression of markers in R1 cells



$CD19^+$ B cells in these transgenic mice had expressed iCre. In contrast, only low levels of GFP expression were detected in peripheral B cells from hCD2-GFP mice. To determine the stage of B cell development at which expression of iCre was initiated in this lineage, bone marrow was isolated and stained with anti-CD19 and anti-IgD antibodies (figure 15 and table 2 and 3). $CD19^+/IgD^+$ bone marrow cells are mature recirculating B cells, whereas $CD19^+/IgD^-$ bone marrow cells represent stages of B cell development prior to migration into the periphery. All mature $CD19^+/IgD^+$ and $CD19^+/IgD^-$ bone marrow derived B cells in Vav-iCre/EYFP and hCD2-iCre/EYFP transgenic mice were positive for EYFP expression, indicating that even the earliest B cell progenitors had undergone iCre mediated recombination. In contrast, recirculating $CD19^+/IgD^+$ bone marrow B cells from hCD2-GFP mice expressed lower levels of GFP, comparable to those found in the peripheral lymph node B cells of these mice. Interestingly, the more immature $CD19^+/IgD^-$ bone marrow B cells expressed high levels of GFP. These data suggest that the Vav-iCre and hCD2-iCre constructs have been active in all B cell and T cell progenitors. The pattern of expression of hCD2-GFP that does not detect the history of expression, but current expression, indicates that the transgenic construct is active in early stages of B cell development, but its expression is diminished as the B cells mature leading to a picture resembling PEV in this cell lineage.

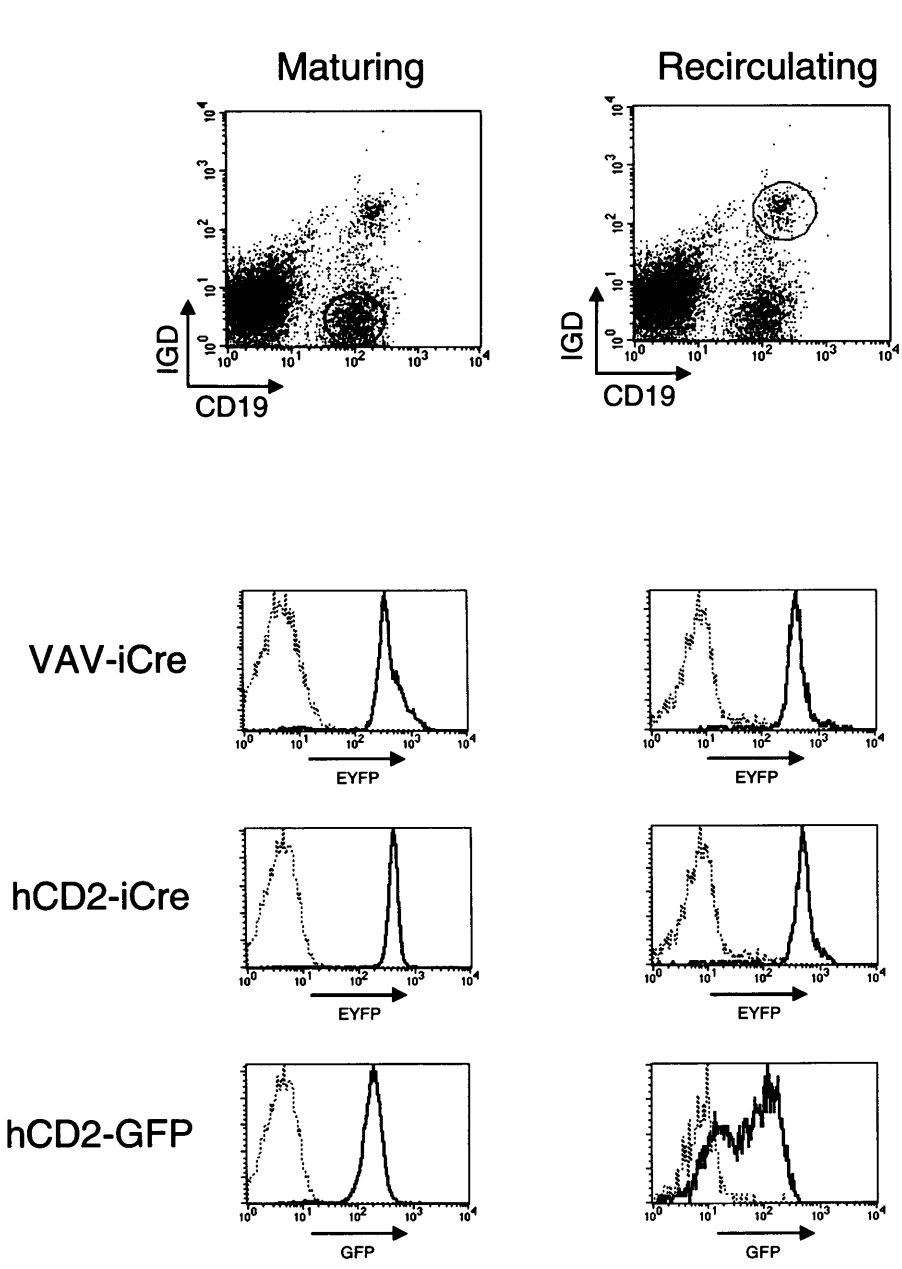
Cre expression in other lineages

As the Vav1 regulatory elements have previously been shown to direct transgene expression in all haematopoietic cells we decided to check for Cre expression in the myeloid and erythroid lineages.

Figure 15 Flow cytometric analysis of EYFP and GFP expression in bone marrow derived B cells

Single cell suspensions from bone marrow from Vav-iCre R26R-EYFP double transgenic mice, hCD2-iCre R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were stained with CD19 and IgD antibodies. Levels of EYFP or GFP expression on mature recirculating CD19+/IgD+ B cells and CD19+/IgD- immature B cells are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the auto-fluorescence in non transgenic controls. Each experiment was repeated at least three times with similar results.

Bone marrow


B cells
Maturing
Recirculating
IGD
CD19
IGD
CD19
VAV-iCre
EYFP
EYFP
hCD2-iCre
EYFP
EYFP
hCD2-GFP
GFP
GFP

Firstly, bone marrow cells were stained with anti-Gr-1 and anti-Mac-1 to identify granulocytes and monocytes (figure 16 and table 2 and 3). Granulocytes ($Gr\text{-}1^{+}/Mac\text{-}1^{+}$) from the bone marrow of Vav-iCre/EYFP transgenic mice showed high levels of EYFP expression in all cells. In contrast, no expression of EYFP or GFP was seen in granulocytes of hCD2-iCre/EYFP or hCD2-GFP transgenic mice, respectively. All monocytes ($Gr\text{-}1^{+}/Mac\text{-}1^{low}$) from Vav-iCre/EYFP transgenic mice showed high levels of EYFP. No substantial EYFP or GFP expression was seen in monocytes from either hCD2-iCre/EYFP or hCD2-GFP transgenic mice, respectively. However, a small population of EYFP positive $Gr\text{-}1^{+}/Mac\text{-}1^{low}$ cells were observed in hCD2-iCre/EYFP transgenic mice.

Finally, to analyse EYFP expression in the erythroid lineage, bone marrow was stained with anti-Ter119 (figure 16 and table 3). Erythrocytes from Vav-iCre/EYFP transgenic mice were seen to express EYFP, although the levels of expression varied significantly between cells. No EYFP or GFP expression was seen in erythrocytes from hCD2-iCre/EYFP or hCD2-GFP transgenic mice, respectively.

Figure 16 Flow cytometric analysis of EYFP and GFP expression in bone marrow derived myeloid and erythroid cells

Single cell suspensions from bone marrow from Vav-iCre R26R-EYFP double transgenic mice, hCD2-iCre R26R-EYFP double transgenic mice, hCD2-GFP mice and wild type mice were stained with Gr-1 and Mac-1 or Ter119 antibodies. Levels of EYFP or GFP expression on granulocytes Gr-1^{+}/Mac-1^{+}, monocytes Gr-1^{+}/Mac-1^{low} or erythrocytes are shown as histograms. Bold line indicates EYFP or GFP expression in the transgenic animals while the dotted line shows the auto-fluorescence in non transgenic controls. Each experiment was repeated at least three times with similar results.

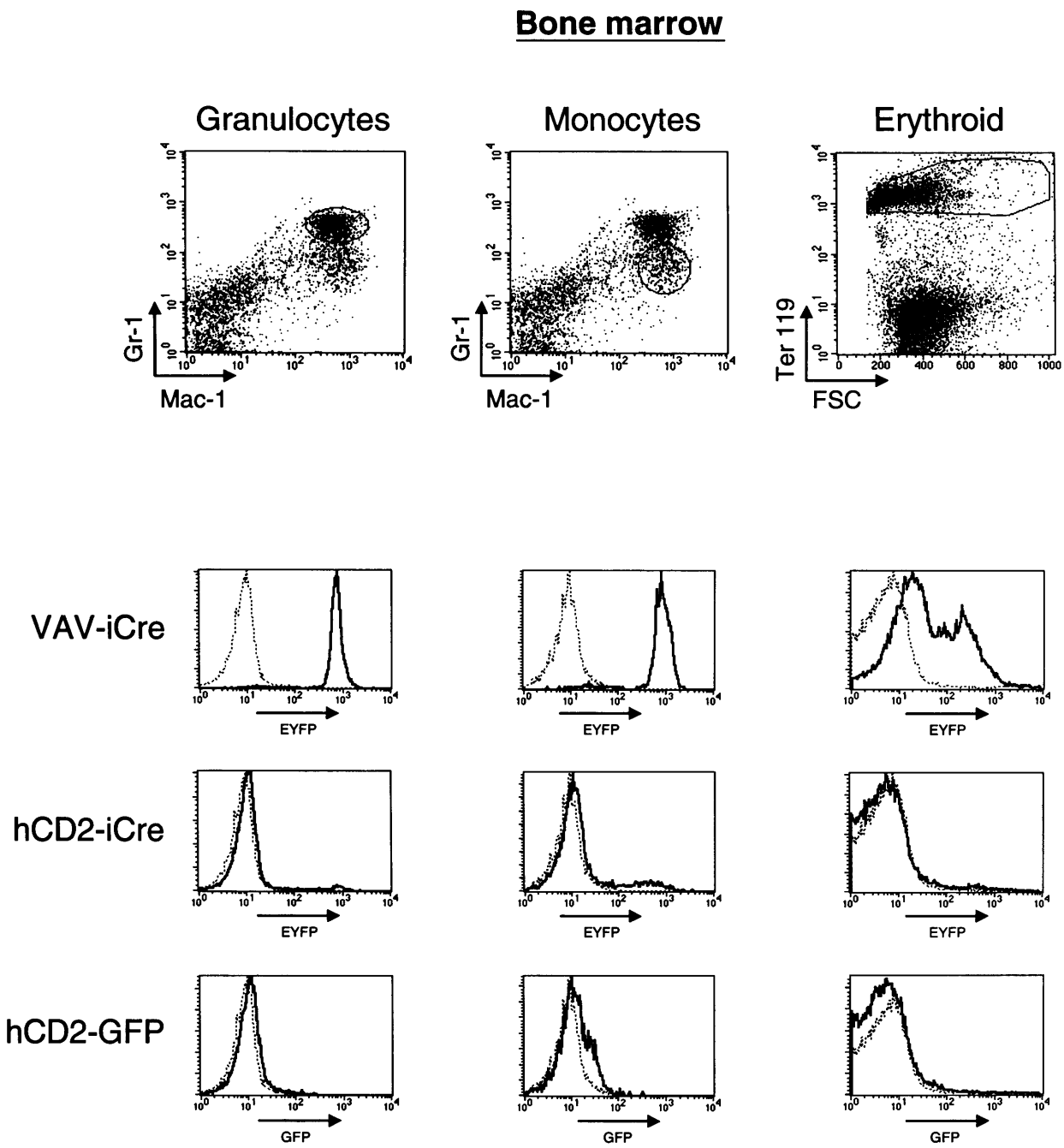

Bone marrow
Granulocytes
Monocytes
Erythroid
Gr-1
Mac-1
Gr-1
Mac-1
Ter 119
FSC
VAV-iCre
hCD2-iCre
hCD2-GFP
EYFP
EYFP
EYFP
EYFP
EYFP
EYFP
GFP
GFP
GFP

Table 2: Summary of iCre mice: Cellularity

The table shows the Absolute number of cells (ABS) and the percentage of cells for different cell types in haematopoietic organs. The mean of triplicates ± one standard deviation is shown.

	Vav-iCre		hCD2-iCre		hCD2-GFP		Non transgenic	
	ABS (x1,000,000)	%	ABS (x1,000,000)	%	ABS (x1,000,000)	%	ABS (x1,000,000)	%
Lymph node*								
Peripheral T cells	6.1±2.6	45%±3	7.1±3.6	45%±3	4.2±3.0	46%±3	4.5±3.3	44%±3
Peripheral B cells	5.4±2.3	40%±3	6.5±3.3	41%±3	3.5±2.5	38%±3	3.8±3.0	40%±3
Thymus								
Double Negative	6.2±0.7	4%±1	4.9±0.4	3%±1	4.3±0.4	4%±1	6.6±0.4	5%±1
Double Positive	130±15	84%±1	140±9.9	85%±1	93±9.0	86%±1	110±6.2	84%±1
Single CD4 Pos.	14±1.6	9%±1	16±1.2	10%±1	9.8±1.0	9%±1	12±0.7	9%±1
Single CD8 Pos.	4.7±0.5	3%±1	4.9±3.5	3%±1	4.3±0.5	4%±1	4.0±0.2	3%±1
Double Negative Stages								
DN1	N.D.	20%±2	N.D.	22%±2	N.D.	19%±2	N.D.	20%±2
DN2	N.D.	6%±3	N.D.	5%±3	N.D.	6%±3	N.D.	8%±3
DN3	N.D.	43%±2	N.D.	42%±2	N.D.	41%±2	N.D.	44%±2
DN4	N.D.	31%±3	N.D.	33%±3	N.D.	30%±3	N.D.	32%±3
Bone marrow								
Maturing B cells	1.0±0.31	20%±4	1.0±0.51	24%±4	0.76± 0.15	20%±4	0.82±0.29	19%±4
Recirculating B cells	0.1±0.02	2%±1	0.09±0.03	2%±1	0.08±0.02	2%±1	0.09±0.03	2%±1
Granulocytes	1.2±0.32	25%±4	1.3±0.54	30%±4	0.80±0.15	21%±4	0.95±0.34	22%±4
Monocytes	0.25±0.06	5%±2	0.17±0.07	4%±2	0.19±0.04	5%±2	0.22±0.08	5%±2
Erythroid	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

* The cells are isolated from the following lymph nodes: Axillary, Brachial, Mesenteric and Inguinal lymph nodes.

** Cells are isolated from the bones of one back leg.

N.D. Not Done

Table 3: Summary of iCre mice: Reporter expression in analysed cells.

The table shows the percentage of EYFP positive cells for the analysed cell types. The mean of triplicates ± one standard deviation is shown.

	Vav-iCre	hCD2-iCre	hCD2-GFP	Non transgenic
	% EYFP Positive	% EYFP Positive	% EYFP Positive	% EYFP Positive
Lymph node				
Peripheral T cells	99%±1	98%±2	98%±2	0%
Peripheral B cells	99%±1	99%±1	Variegated	0%
Thymus				
Double Negative	97%±1	93%±1	90%±1	0%
Double Positive	100%±1	99%±1	99%±1	0%
Single CD4 Pos.	99%±1	99%±1	98%±1	0%
Single CD8 Pos.	98%±1	97%±2	95%±2	0%
Double Negative Stages				
DN1	100%±1	99%±1	43%±3	0%
DN2	93%±2	41%±21	52%±4	0%
DN3	99%±1	98%±1	98%±1	0%
DN4	100%±1	99%±1	98%±1	0%
Bone marrow				
Maturing B cells	99%±1	100%±1	99%±1	0%
Recirculating B cells	99%±1	99%±1	65%±4	0%
Granulocytes	99%±1	1%±0	0%	0%
Monocytes	99%±1	2%±1	0%	0%
Erythroid	Variegated	0%	0%	0%

Generation of mice with targeted HBP1 allele.

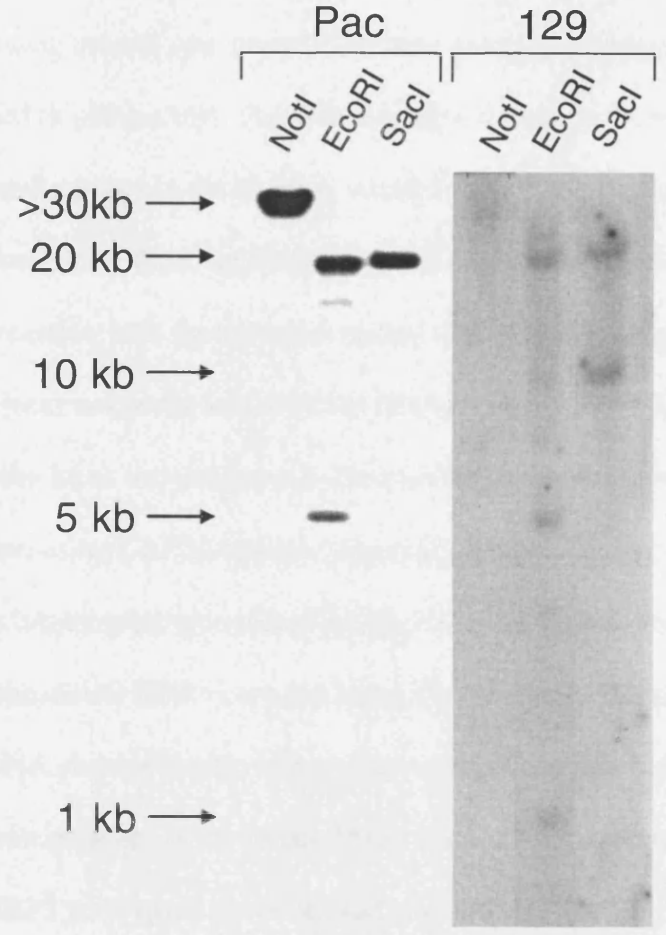
Screening of mouse genomic PAC library RPCI21

The first step in generating mice with a targeted HBP1 allele was to obtain a genomic clone, which contains the HBP1 locus. Previously in our laboratory a mouse genomic phage library was screened to obtain the HBP1 genomic locus. Initial mapping of these clones let to the identification of two types of clones (data not shown). Clones containing the HBP1 gene as one fragment without an intron and exon structure and clones with an intron and exon structure. Intronless copies of genes are termed processed pseudo genes and are non-functional. Processed pseudo genes are created through a process called retro-transposition during this process mRNA transcript gets reverse transcribed with subsequent re-integration of the cDNA into the genome (Zhang et al., 2004). The clones containing the pseudo gene were not further characterised. The isolated clones contained only part of the HBP1 locus. In order to obtain the entire HBP1 locus a PAC library was screened. The mouse PAC library RPCI21 was provided by the MRC gene service and was constructed by Kazutoyo Osoegawa (Osoegawa et al., 2000). The source for the library is mouse spleen genomic DNA from a female 129/SvevTACfBr. The average insert size is 147 kb and the library consists of approx. 123,000 clones. The library has been gridded on Hybond N nylon membranes (Amersham). The filters were hybridized with a 2.4 kb genomic fragment of HBP1 isolated from a mouse genomic phage library (Stratagene) that should not cross hybridize with the pseudo-gene. Eleven positive clones were

obtained from the Resource Centre. The analysis of one of these clones, N621-G6 is briefly described. The DNA of clone N621-G6 was digested with NotI, EcoRI and SacI and analysed by southern analysis. Figure 17 shows the hybridization pattern using the open reading frame of the HBP1 cDNA. The open reading frame was used as a probe to ensure that the entire HBP1 locus will be present in the PAC clone. The drawback of using this probe is the high sequence similarity with the pseudo-gene. Pseudo-genes often interfere with PCR or hybridization experiments that are intended for the "real" genes. The hybridisation fragments are identical in size between clone N621-G6 and genomic mouse DNA, although the genomic DNA shows extra bands (extra 1kb band with EcoRI digest and extra 10kb band with SacI digest) corresponding to the pseudo gene. Because the bands are identical in size between the PAC and genomic DNA we conclude that clone N621-G6 encompasses the entire HBP1 genomic locus.

Figure 17 Southern blot analysis of Pac clone N621-G6

DNA from PAC clone N621-G6 and mouse genomic DNA (129/Sv origin) was digested with NotI, EcoRI or SacI respectively. The blot is probed with the open reading frame of HBP1.

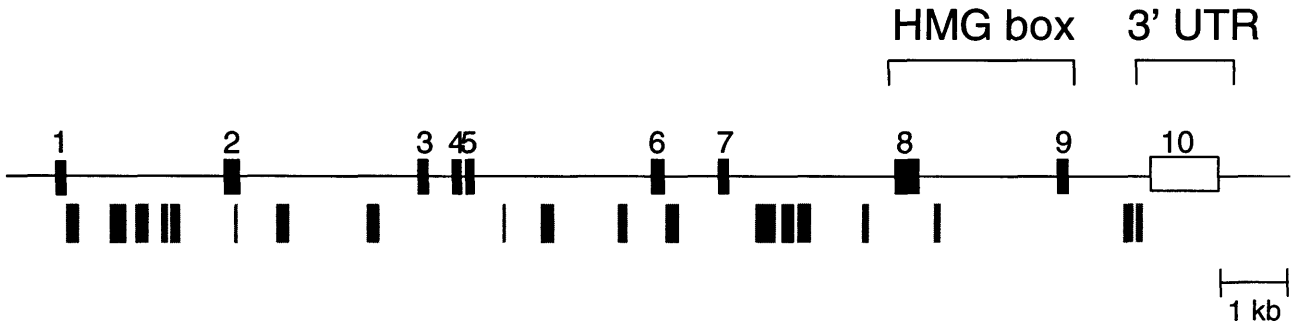

Pac
129
NotI
EcoRI
SacI
NotI
EcoRI
SacI
>30kb
20 kb
10 kb
5 kb
1 kb

Sequencing and mapping of HBP1 locus

To design a targeting vector and to gain insights in the genomic structure of the mouse HBP1 locus a sequencing project was started. Different genomic fragments of the HBP1 were sub cloned in pBluescript. The cloned inserts were sequenced from "universal" primer binding sites in the flanking vector DNA. The next step was to design a walking primer using the most distant, reliable sequence data obtained from the first sequencing reaction with the universal primer. This walking primer was then used to sequence the next unknown section of the DNA template. The last step was repeated until the entire locus was sequenced. The resulting sequence information was compiled by computer, using CAP3 Sequence Assembly Program (program available at http://fenice.tigem.it/bioprg/interfaces/cap3.html), into a contiguous sequence in order to reassemble the mouse HBP1 genomic locus. By comparing the sequence of our mouse HBP1 cDNA clone with that of the genomic sequence data we were able to define the intron / exon structure of the mouse HBP1 gene, which is shown in figure 18. The mRNA of HBP1 gene spans ten exons and nine introns. The HBP1 protein is encoded by exons 1–9 plus a small segment at the 5'-end of exon 10. Most of exon 10, which is the largest exon, does not code for protein sequences (figure 18). The generated map of the mouse HBP1 locus was used to design the targeting vector.

Figure 18 Overview of HBP1 locus

Intron-exon organization of the mouse HBP1 gene: black boxes correspond to coding exons and white boxes represent non-coding exons. The grey boxes indicate the repeat elements as identified by RepeatMasker.


HMG box
3' UTR
1
2
3
4
5
6
7
8
9
10
1 kb

Determination of the location of repetitive DNA

Gene targeting vectors require long regions of target homology and will often include intron sequences or other untranslated segments of the endogenous gene. Untranslated portions of genes can be problematic for the constructions of a targeting vector. These regions can contain stretches of interspersed repetitive DNA that, if located within several hundred base pairs of linearised end of the vector, will drive the vector to target homologous repetitive sequences in the genome. This leads to homologous recombination in other places than the target gene. To determine any major stretches of repetitive DNA the sequenced fragments from the mouse HBP1 locus were analysed using RepeatMasker. RepeatMasker is a program that screens DNA sequences for interspersed repeats known to exist in mammalian genomes as well as for low complexity DNA sequences (program available at: http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). The mouse HBP1 locus contains about 20% of repetitive sequences (figure 18). The longest identified repeat is an Alu repeat of 153bp. The targeting vector was designed so that at least 0.5 kb of non-repetitive DNA was present at the 5' and 3' end of the homologous DNA sequence.

Generation of mice with targeted allele.

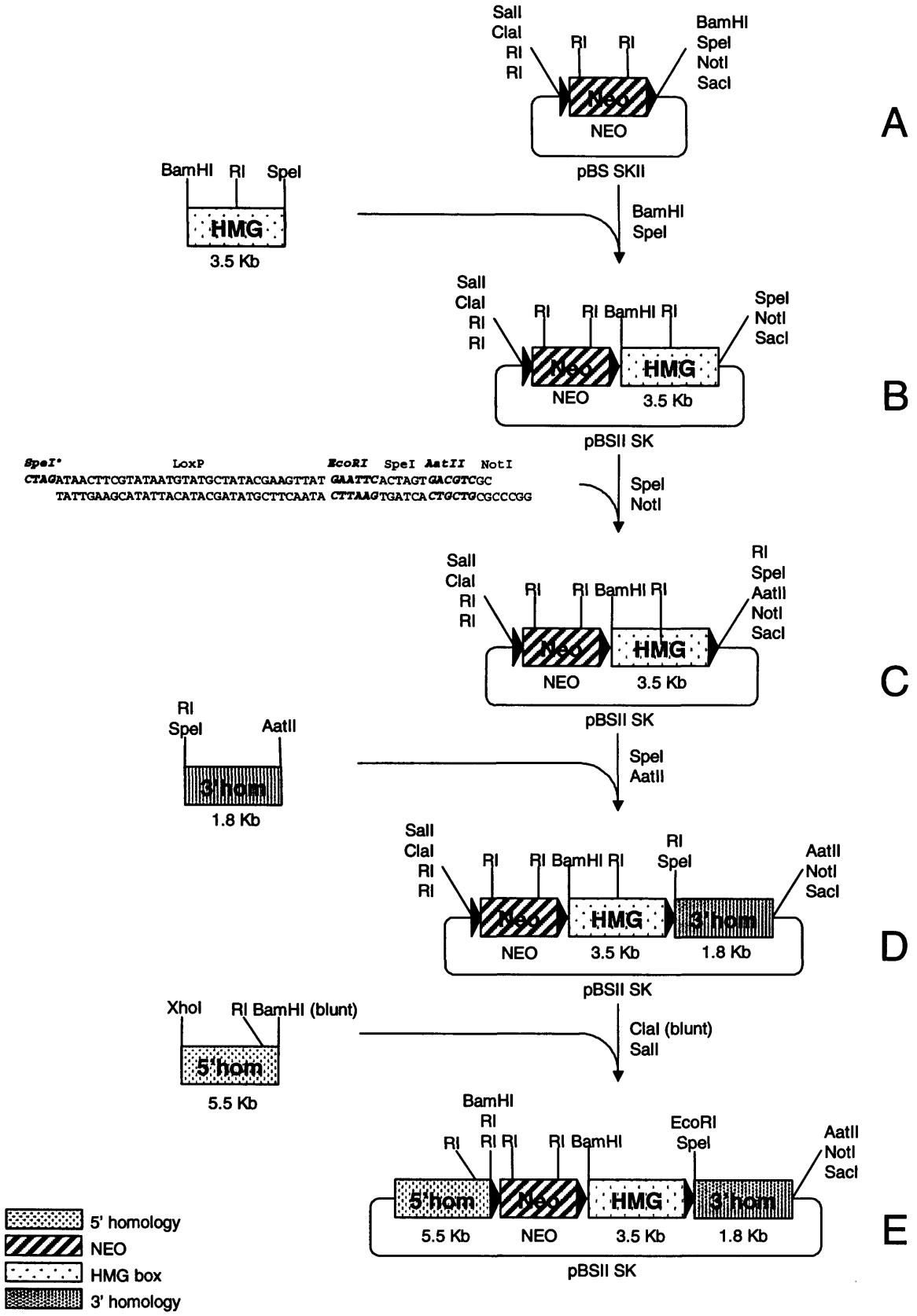
We chose to flox the exons 8 and 9 of the HBP1 locus by homologous recombination because they encode the HMG box, which provides the DNA binding domain for the HBP1 protein. We reasoned that without a DNA binding domain some functions of HBP1 will be abolished. Furthermore, in the nuclear localisation studies described above we had shown that the HMG box contains important sequences for translocation of the protein into the nucleus.

In order to flox the HMG box of HBP1 in the mouse genome, a targeting construct with floxed HMG box encoding exons was introduced into embryonic stem cells.

In order to generate the targeting construct (figure 19) the following steps were taken; pBSneo, which contains a loxP-flanked neomycin gene cloned into the SmaI site of pBluescript II SK, was digested with BamHI and SpeI. A 3.5 kb BamHI SpeI genomic fragment containing the HMG box encoding exons was inserted in the BamHI SpeI digested pBSneo to generate construct B (figure 19). A double stranded oligo nucleotide containing a loxP site and an EcoRI site was synthesised and cloned into the SpeI and NotI site of construct B to generate construct C (figure 19). Construct C was cut open with SpeI and AatII and a 1.8 kb SpeI-AatII genomic fragment consisting of 3' homology was cloned into SpeI AatII digested construct C to generate construct D (figure 19). Finally, construct D was digested with ClaI, blunted and then cut with SalI. A 5.8 kb XhoI-BamHI (blunt) genomic fragment consisting of the 5'homology of the targeting construct was cloned into SalI ClaI (blunt) to generate the final targeting construct (figure 19). The target construct was designed so that when

Figure 19 Schematic representation of the cloning strategy

See text for details. The diagram is not to scale.


SalI
ClaI
RI
RI
RI
RI
BamHI
SpeI
NotI
SacI
Neo
NEO
A
pBS SKII
BamHI RI SpeI
HMG
3.5 Kb
BamHI
SpeI
SalI
ClaI
RI
RI
RI RI BamHI RI
SpeI
NotI
SacI
Neo
HMG
NEO
3.5 Kb
B
pBSII SK
SpeI* LoxP EcoRI SpeI AatII NotI
CTAGATAACTTCGTATAATGTATGCTATACGAAGTTAT GAATTCACTAGTGACGTCGC
TATTGAAGCATATTACATACGATATGCTTCAATA CTTAAGTGATCACTGCTGCGCCCGG
SpeI
NotI
SalI
ClaI
RI
RI
RI RI BamHI RI
RI
SpeI
AatII
NotI
SacI
Neo
HMG
NEO
3.5 Kb
C
pBSII SK
RI
SpeI
AatII
3'hom
1.8 Kb
SpeI
AatII
SalI
ClaI
RI
RI
RI RI BamHI RI
RI
SpeI
AatII
NotI
SacI
Neo
HMG
3'hom
NEO
3.5 Kb
1.8 Kb
D
pBSII SK
XhoI
RI BamHI (blunt)
5'hom
5.5 Kb
ClaI (blunt)
SalI
BamHI
RI
RI
RI RI
RI BamHI
EcoRI
SpeI
AatII
NotI
SacI
5'hom
Neo
HMG
3'hom
5.5 Kb
NEO
3.5 Kb
1.8 Kb
E
pBSII SK
5' homology
NEO
HMG box
3' homology

replacement of the endogenous sequence by the exogenous one occurred via homologous recombination, the HMG box encoding sequences of the HBP1 gene would become floxed. The Neo gene would also be transferred into the targeted site. Thus, cells in which the homologous recombination occurred would be neo resistant.

In order to generate embryonic stem cells carrying one allele of the HBP1 gene with floxed HMG box encoding exons, the construct was linearised with SacI and introduced into PC3 embryonic stem cells of 129/Sv origin (O'Gorman et al., 1997) by electroporation. $0.8x10^7$ ES cells were transfected by electroporation with a final DNA concentration of 45μg/ml. Selection was initiated 24hrs later at a concentration of 250 or 300μg/ml G418. The targeting event inserts the floxed neomycin resistance gene 5' of the HMG box encoding exons and a loxP site 3' of the HMG box encoding exons. The homologous recombination event is shown in figure 20. Because of the artificially introduced EcoRI site in the end of the 3' loxP site, the 4.4 kb wild type EcoRI fragment is converted to a smaller 1.9 kb mutant fragment, as detected by a 400bp NcoI-EcoR1 probe. This probe is located downstream of the targeted region, 3' of the homology of the targeting vector (figure 20). 576 resistant clones were screened by Southern blot analysis, using the 400bp NcoI EcoR1 probe, six clones that showed the 1.9kb mutant band were picked (for an example see figure 20D). On the southern blots the probe gives, besides the 4.4 kb wildtype and the 1.9 kb band of the targeted allele, a 3.5 kb band (* in figure 20). This band probably represents the pseudo gene of HBP1, which is present in the mouse genome. The integrity of the locus was verified with a SacI digest using the Neo probe. The SacI digest with the Neo probe resulted in a 21kb fragment (for an example see figure 20D). Besides an 21 kb

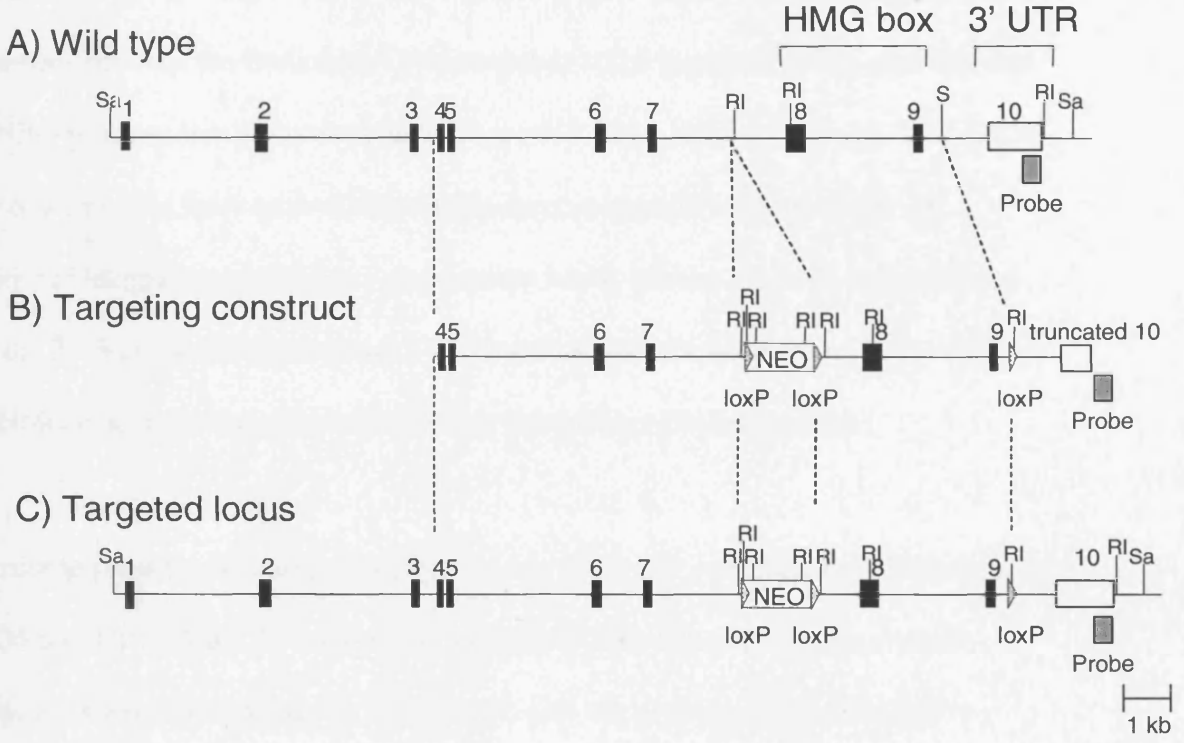
Figure 20 Targeting of HBP1 locus

A) Schematic representation of the mouse HBP1 locus. Map of the HBP1 gene locus showing the location of the HMG box (black boxes: coding exons white box: untranslated region). Restriction enzyme sites are indicated by vertical bars (RI: EcoRI, S: SpeI, Sa: SacI).

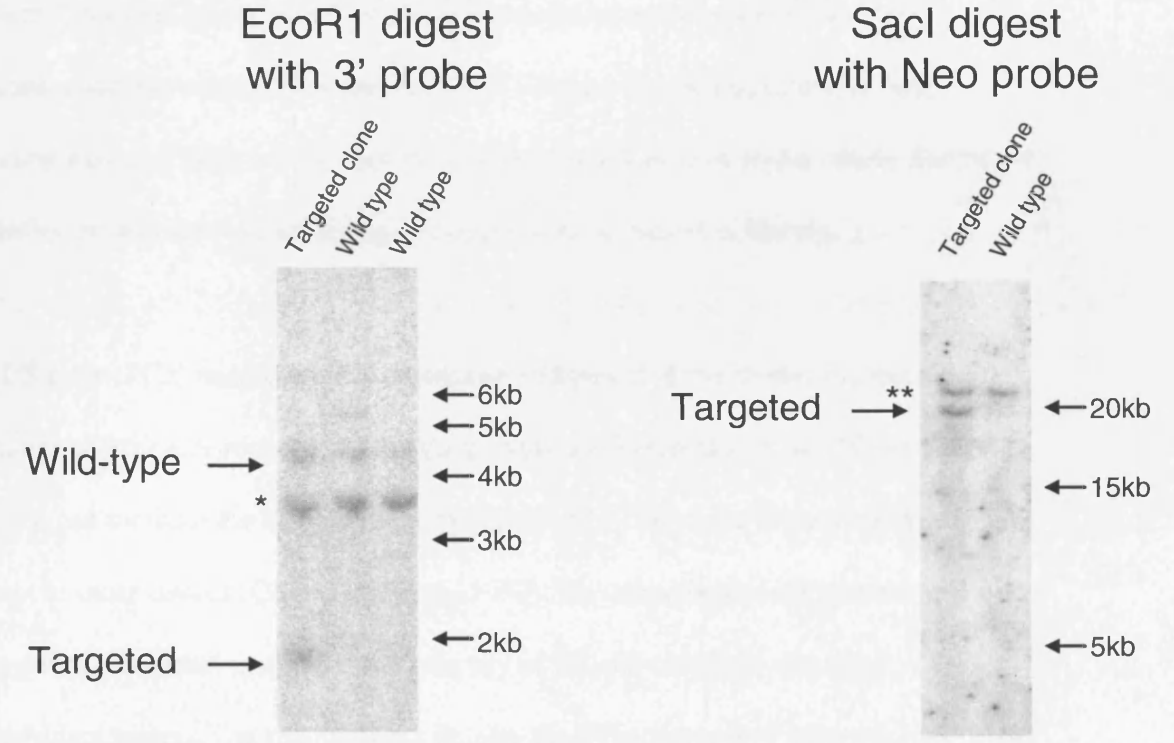
B) The targeting construct. LoxP sites are indicated by grey arrows. The neomycin gene is indicated.

C) The targeted locus. Map of the targeted HBP1 locus after homologues recombination with the targeting vector.

D) Southern analysis of targeted clones. The insertion of the floxed Neomycin and loxP with EcoRI site results in the reduction of a wild-type 4.4 kb EcoRI fragment to a mutant 1.9 kb EcoR1 fragment as detected with a 0.4 kb NcoI EcoRI genomic probe (grey box) as shown in the southern blot analysis. The probe also detects a 3.5kb fragment (indicated with *) corresponding to the pseudo-gene of HBP1. The insertion of the floxed Neomycin and loxP result in a 21 kb SacI fragment. The Southern blot was labelled with the NEO probe. The probe also cross hybridizes with a 21 kb SacI fragment (indicated with **).



D) Southern blot analysis of targeted clones



fragment for the targeted locus the probe used also cross hybridizes with a 23 kb SacI fragment (** in figure 20). The 23 kb fragment is probably due to homologues sequences between the Protamine Cre transgene, which is present in the used ES, and the NEO probe as it is not present in wildtype 129 DNA (data not shown). This digest also confirms that there is no additional random integration in the genome. An additional integration would have shown extra bands besides the 21kb targeted band and the 23kb cross-hybridizing band. These results indicated that ES cells in which the HMG box encoding exons of HBP1 are floxed have been generated.

In order to generate chimeric mice, three of the six clones were karyotyped (number 10, 35 and 126); all the three clones had a euploid karyotype (40 chromosomes). These cells were injected into 3.5-day blastocysts derived from C57Bl/6 females. These injected blastocysts were subsequently transferred into the uteri of foster mothers. Chimeric males were born as determined by agouti coat colour. The chimeric mice that carried the targeted HBP1 allele with the floxed HMG box encoding exons of HBP1 were bred to establish mouse lines with this allele. Germline transmission was determined by agouti coat colour of the F1 offspring.

The ES cells (PC3) used contain a transgene comprised of the mouse Protamine 1 promoter and the Cre recombinase coding sequence (O'Gorman et al., 1997). These ES cells can mediate the efficient recombination of a Cre target in the male germ line, but not in other tissues (O'Gorman et al., 1997). Cre recombinase expression in the male germ line would mean that the progeny of ES cell chimeras can carry recombinant targets. Cre recombinase deletes the DNA sequences between loxP sites when the loxP sites are in the same orientation. Because there were three loxP sites in

the targeted HBP1 allele, there were three possible Cre-mediated recombination events (figure 22): deletion of the neo alone (generating $HBP1^{+/flHMG}$ mice), deletion of the HMG box of the HBP1 gene but not the neo (generating $HBP1^{+/neo}$ mice), and deletion of both the HMG box and the neo (generating $HBP1^{+/\Delta HMG}$ mice). The last possibility is that Cre is not active and no recombination between the loxP sites occurs (generating $HBP1^{+/flHMGneo}$ mice). To establish which of these recombination events has occurred in the progeny of ES cell chimera tail DNA was isolated and analysed by Southern blot or PCR analysis.

Progeny of ES cell chimeras were tested by Southern analysis using EcoRV digestion followed by hybridization with a probe specific for the region between the third loxP sites and the EcoRV recognition sequence (figure 21). A 7.0 kb EcoRV fragment represents the wild type locus. Mice carrying the targeted HBP1 locus can show the following additional fragments depending on which loxP sites recombined. If Cre was inactive in the germline the EcoRV fragment from the targeted locus will be 6kb (figure 21). Deletion of the neo alone (generating $HBP1^{+/flHMG}$ mice) will generate a 4.7 kb fragment. Deletion of the HMG box of the HBP1 gene, but not the neo (generating $HBP1^{+/neo}$ mice), will generate a 2.5 kb fragment. Finally, deletion of both the HMG box and the neo (generating $HBP1^{+/\Delta HMG}$ mice) will generate a 1.2 kb fragment (figure 21). Southern blot analysis of the progeny revealed, in addition to wild type 7.0 kb band, either a 4.7 kb fragment (corresponding to $HBP1^{+/flHMG}$ mice) or a 1.2 kb fragment (corresponding to $HBP1^{+/\Delta HMG}$ mice). For an example of these deletions see figure 22A. Deletion of only the floxed HMG box (without deleting NEO) was not observed in the mice analysed. This indicated that mice with a

Figure 21 Schematic overview of partial Cre recombination

Schematic representations of HBP1, showing the targeted and Cre-mediated recombinant alleles. For consistency with the text and figures, each allele is named. Also shown are schematic representations of the probe (grey) and of the primers (P1-P4) used to determine the HBP1 genotype.

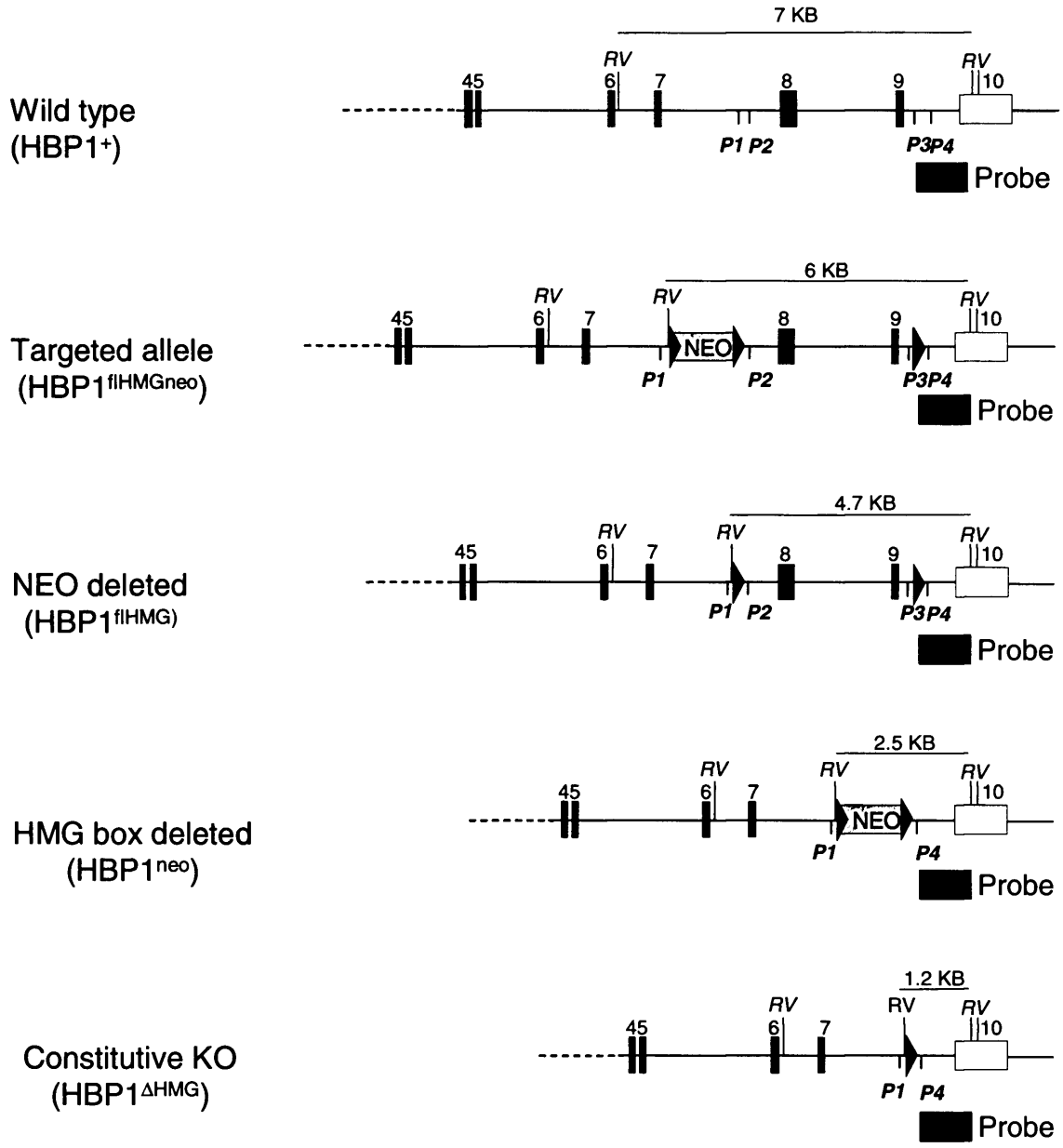
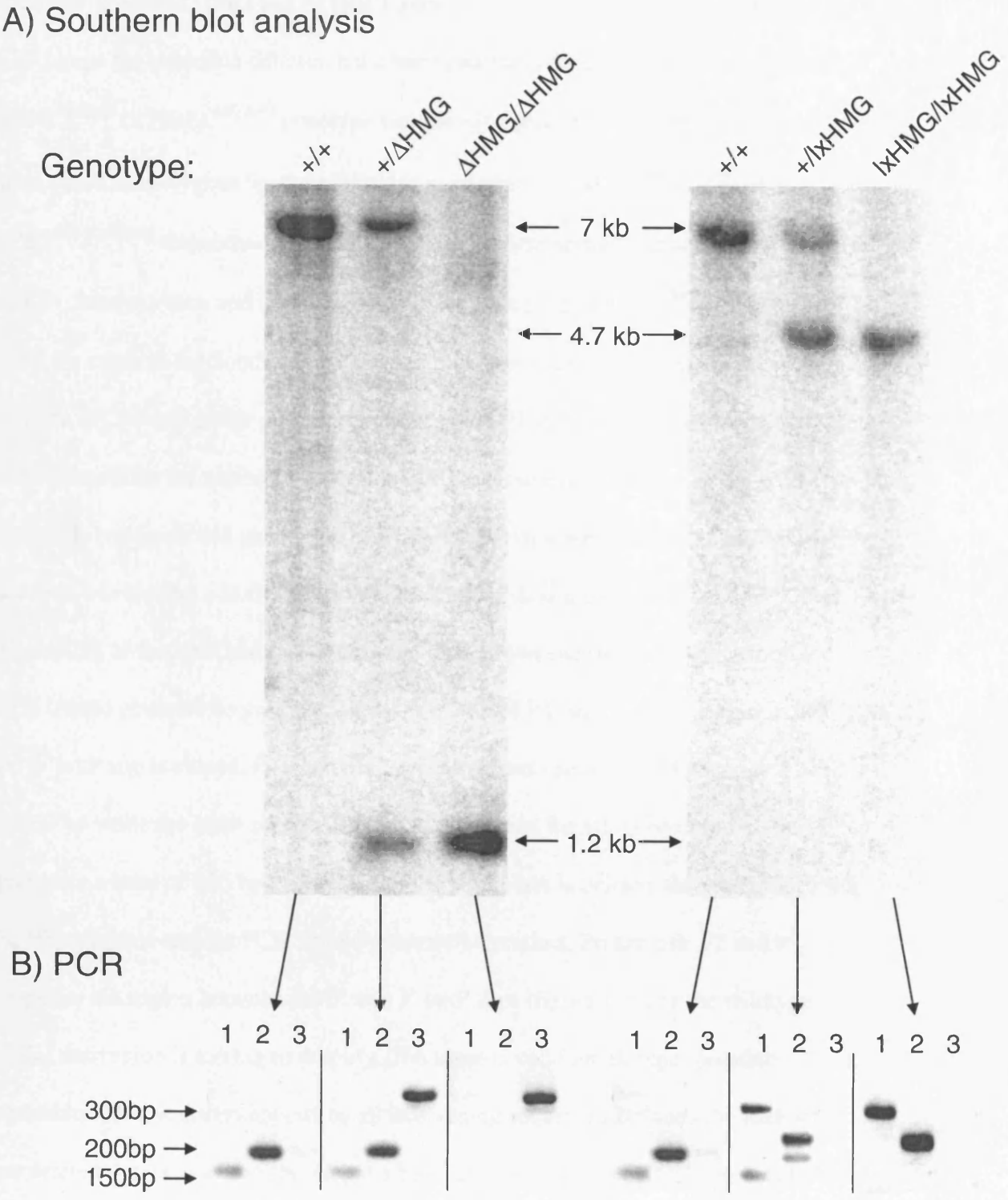

Wild type
(HBP1+)
7 KB
RV
RV
45
6
7
8
9
10
P1 P2
P3P4
Probe
Targeted allele
(HBP1flHMGneo)
6 KB
RV
RV
RV
45
6
7
NEO
8
9
10
P1
P2
P3P4
Probe
NEO deleted
(HBP1flHMG)
4.7 KB
RV
RV
RV
45
6
7
8
9
10
P1
P2
P3P4
Probe
HMG box deleted
(HBP1neo)
2.5 KB
RV
RV
RV
45
6
7
NEO
10
P1
P4
Probe
Constitutive KO
(HBP1ΔHMG)
1.2 KB
RV
RV
RV
45
6
7
10
P1
P4
Probe

Figure 22 Cre-mediated recombination of the HBP1 allele

A) Southern blot analysis was performed on tail DNA from wildtype, heterozygous and homozygous mice. The probe is indicated in figure 21.

B) The Southern blots were confirmed by PCR analysis. The primers are indicated in figure 21. Primer pair P1 and P2 is shown in lane 1, primer pair P3 and P4 is shown in lane two, and finally primer pair P1 and P4 is shown in lane three.


A) Southern blot analysis
Genotype:
+/+
+/ΔHMG
ΔHMG/ΔHMG
+/+
+/lxHMG
lxHMG/lxHMG
7 kb
4.7 kb
1.2 kb
B) PCR
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
300bp
200bp
150bp

constitutive deleted HMG box of HBP1 gene and mice with a floxed HMG box of the HBP1 gene for inducible deletion have been generated. Heterozygous mice with the $HBP1^{+/\Delta HMG}$ or $HBP1^{+/flHMG}$ genotype were bred to generate the homozygous mutant mice. Mice homozygous for the mutations are designated $HBP1^{\Delta HMG/\Delta HMG}$ or $HBP1^{flHMG/flHMG}$ respectively. Figure 22 shows the Southern blot analysis of wild type 129/Sv, heterozygous and homozygous mutant mice of the $HBP1^{\Delta HMG}$ and $HBP1^{flHMG}$ mice. To confirm the Southern analysis the mice were screened by polymerase chain reaction (PCR) using four different primers called P1 to P4 (figure 21) Primer pair P1 and P2 amplifies the region in which the NEO is cloned (figure 21). For the wild-type allele this region should generate a band of 170bp, while the same region after Neo deletion, but leaving a loxP site and the HMG box, should generate a band of 300bp (figure 21). If the HMG box is deleted, the recognition site for P2 is deleted and the PCR should generate no product. Primer pair P3 and P4 amplifies the region in which the 3' loxP site is cloned. For the wild-type allele, this region should generate a band of 200 bp while the same region with the loxP site and the HMG box not deleted generates a band of 250 bp (figure 21). If the HMG box is deleted, the recognition site for P3 is deleted and the PCR should generate no product. Primer pair P1 and P4 amplifies the region between the 5' and 3' loxP sites (figure 21). For the wildtype allele, this region is too big to amplify. The same is valid for all other possible recombination events excepts when all intervening sequences between the loxP sites are deleted in this case you should get a band of 350 bp (figure 21). Using the above PCR strategy the southern blots were confirmed as shown in figure 22B.

Northern Analysis

To confirm that the HMG box was removed from the HBP1 gene mRNA from wildtype and knockout was analyzed by Northern analysis. Northern analysis provides a direct relative comparison of message abundance and transcript size between samples on a single membrane. RNA was isolated from the thymus of $HBP1^{+/+}$, $HBP1^{+/\Delta HMG}$ and $HBP1^{\Delta HMG/\Delta HMG}$ mice. RNA samples were separated by size via electrophoresis in an agarose gel under denaturing conditions. The RNA was transferred to an N+ membrane, cross-linked and hybridized with a labelled probe. Northern hybridization was performed with radiolabelled open reading frame cDNA of HBP1.

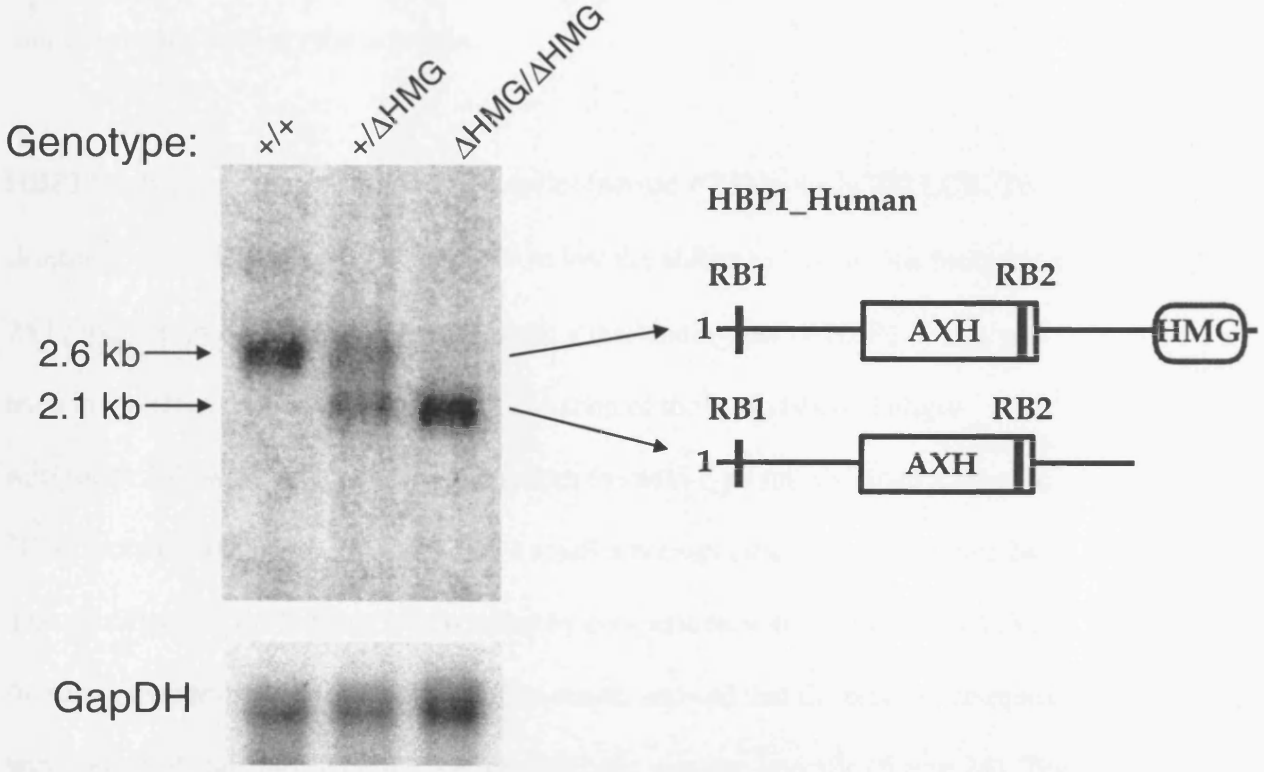
Figure 23 presents such northern blot analysis. Full length HBP1 was detected in $HBP1^{+/+}$ mice as a single transcript of approximately 2.6 kb in size. In the $HBP1^{+/\Delta HMG}$ mice, a full length HBP1 was detected as well as a smaller transcript of 2.1 kb in size that corresponds to HBP1 without the HMG box ($HBP1^{\Delta HMG}$). In the last lane, RNA from a mouse homozygous for the ΔHMG mutation shows a single transcript corresponding to HBP1 without the HMG box ($HBP1^{\Delta HMG}$). This confirms that the HMG box from HBP1 has been deleted.

The DNA mobility shift assay

The HMG box is a conserved DNA-binding domain found in a family of transcription factors that regulate growth and development. Therefore, we postulated that

Figure 23 Northern-blot analysis of mRNA from HBP1$^{\Delta HMG}$ mice

mRNA from wildtype, heterozygous and homozygous HBP1$^{\Delta HMG}$ mice was probed with the HBP1 coding sequences. A 2.6 kb transcript was detected in the wildtype and heterozygous HBP1$^{\Delta HMG}$ mice. A smaller transcript of 2.1 kb, corresponding to HBP1 lacking the HMG box, was detected in heterozygous and homozygous HBP1$^{\Delta HMG}$ mice. The membrane was subsequently incubated with GapDH to compare relative amounts of mRNA in each lane.


Genotype:
+/+
+/ΔHMG
ΔHMG/ΔHMG
2.6 kb
2.1 kb
GapDH
HBP1_Human
RB1
RB2
1
AXH
HMG
RB1
RB2
1
AXH

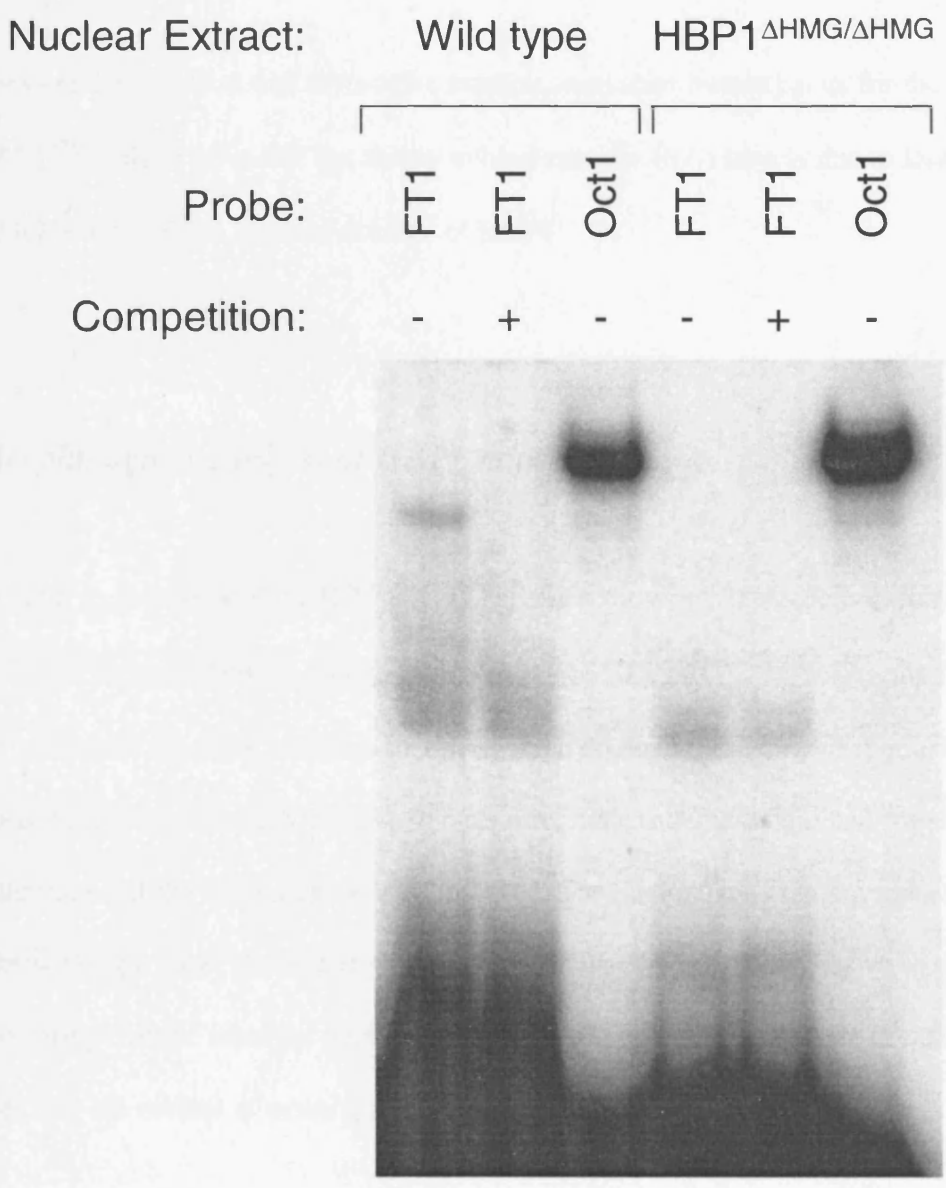
sequences within the HMG box must be required to maintain the DNA binding capacities of the HBP1 protein. The DNA mobility shift assay provides a tool for the detection of factors binding to specific sequences. The method relies on the ability of a protein to bind to radiolabelled DNA fragment (probe) in vitro. This is followed by electrophoretic separation of DNA:protein complexes from the unbound DNA on non-denaturing poly-acrylamide gels.

HBP1 has been reported to bind to a footprint (named FT1) in the hCD2 LCR. To determine if the HBP1$^{\Delta HMG/\Delta HMG}$ mice have lost the ability to bind to this footprint a 24 bp synthetic oligo-nucleotide, representing the binding site of HBP1 (FT1), was used in the DNA mobility shift assay. Incubation of the radio labelled oligo-nucleotide FT1 with the nuclear extracts from the wild-type mice yielded a specific DNA:protein binding complex and also a smaller non-specific complex (figure 24). The specificity of the binding was verified by competition with the same unlabeled oligo-nucleotides at 100 molar excess. The results showed that the binding complex was abolished and confirmed that the smaller band was non-specific (figure 24). The integrity of the thymus nuclear extracts was confirmed by using OCT1 probe. Incubation of the radio labelled oligo-nucleotide OCT-1 with the nuclear extracts from the wild-type mice yielded a specific DNA:protein binding complex (figure 24).

Incubation of the radio labelled oligo-nucleotide FT1 with the nuclear extracts from the HBP1$^{\Delta HMG/\Delta HMG}$ mice yielded no specific DNA:protein binding complex, only the smaller non-specific complex was detected (figure 24). The integrity of the thymus nuclear extracts was confirmed by using OCT1 probe. Incubation of the oligo-nucleotide OCT-1 with the nuclear extracts from the HBP1$^{\Delta HMG/\Delta HMG}$ mice

Figure 24 Electromobility shift assay (EMSA)

FT1, the radiolabelled specific oligo-nucleotide, was incubated with thymus nuclear extract from wild-type mice or HBP1$^{\Delta HMG/\Delta HMG}$ (lane 1 and 4 respectively). 100-fold molar excess of unlabeled specific probe was used for competition with labelled oligo-nucleotide (lane 2 and 5). Lanes 3 and 6, show OCT1, the radiolabelled specific oligo-nucleotide, incubated with thymus nuclear extract from wild-type mice or HBP1$^{\Delta HMG/\Delta HMG}$ respectively.


Nuclear Extract:
Wild type
HBP1ΔHMG/ΔHMG
Probe:
FT1
FT1
Oct1
FT1
FT1
Oct1
Competition:
-
+
-
-
+
-

yielded the normal specific DNA:protein binding complex (figure 24), indicating that the absence of the specific FT1 complex using HBP1$^{\Delta HMG/\Delta HMG}$ protein extracts is not due to protein degradation in the different samples.

These results confirm that thymocyte extracts from mice homozygous for the HBP1$^{\Delta HMG}$ allele have lost the ability to bind specific DNA targets due to loss of the HMG box, the DNA binding domain of HBP1.

Morphological analysis of HBP1 mice

In order to determine if the HBP1$^{\Delta HMG/\Delta HMG}$ mice show any gross defects they were followed through their development. Recovery of HBP1$^{\Delta HMG/\Delta HMG}$ offspring from the heterozygous breeding indicates that these mice do not exhibit an embryonic lethal phenotype. No obvious difference in litter size, male to female ratio and Mendelian inheritance of the mutant allele was observed. The homozygous mutant mice thrive like wild type mice. Even after weaning (after 3 weeks) no differences were observed indicating that the mice eat normally. From the above observations, we concluded that the mice are normal at weaning.

The next step was to see if they are reproductively normal. A male and female HBP1$^{\Delta HMG/\Delta HMG}$ were taken and set up for breeding. Normal size litters from these breedings were generated indicating that their reproductive capacity remains intact.

Gross necropsy was performed which included an initial physical examination of external surfaces and all orifices as well as an internal examination of tissues and organs in situ. The following were examined: the external surfaces of the brain and spinal column, the nose and neck, with associated organs and tissues; the thoracic, abdominal, and pelvic cavities, with associated organs and tissues; and the musculoskeletal carcass. No obvious differences between wild type and $HBP1^{\Delta HMG/\Delta HMG}$ mice were observed.

Thus, as far as we can determine from these experiments we conclude that the development of $HBP1^{\Delta HMG/\Delta HMG}$ is normal and no gross abnormalities were detected.

Haematological analysis of HBP1 mice

Haematopoiesis requires continuous cell proliferation, lineage commitment, differentiation, and maturation. Each step of haematopoietic development has been characterised by the expression of cell surface markers. HBP1 mediates a link between the cell cycle control machinery and cell differentiation signals. Furthermore, K.M. Lin et.al (Lin et al., 2001a) suggested that HBP1 is actively involved in the differentiation process during haematopoiesis. Therefore, the haematopoietic system provides an excellent model for studying the function of HBP1 in cell proliferation, lineage commitment, differentiation, and maturation.

To determine whether the loss of HBP1 had any effect on haematopoiesis, we analyzed primary and secondary lymphoid organs of wild-type, heterozygous and

knockout mice. There was no obvious reduction or increase in the size of the thymus, spleen, and lymph nodes, and comparable cellularity was exhibited in $HBP1^{+/+}$, $HBP1^{+/\Delta HMG}$ and $HBP1^{\Delta HMG/\Delta HMG}$ littermates. Mice from 4 to 12 weeks old (at least three mice per time point) were analysed using flow cytometrical analysis. The flow cytometric analysis of mice of 6 weeks is shown in figure 25 as an example of the mice analysed and a summary of the analysis is given in table 4. Flow cytometry analysis of developing thymocytes revealed no difference in numbers and proportions of double-negative (CD4- CD8-), double-positive (CD4+ CD8+), and single-positive (CD4+ CD8- or CD4- CD8+) subsets (figure 25A), indicating that T-cell development is unaffected by HBP1 deficiency. Peripheral T-cell numbers, as measured by anti-Thy.1 reactivity, were likewise unaffected by the absence of HBP1, as were the levels and ratio of CD4 or CD8 single-positive T cells (figure 25B and C). With regard to the B cell lineages we observed normal numbers of maturing B cells (CD19+ IgD-), and mature B (CD19+ IgD+) cells in the bone marrow of HBP1 deficient mice (figure 25D). To assess whether the deletion of HBP1 affected mature B-cell development, we analyzed B-cell populations in the lymph nodes by flow cytometry using antibodies to CD19. There was no significant difference in numbers of peripheral B cells between $HBP1^{+/+}$ and $HBP1^{\Delta HMG/\Delta HMG}$ mice (figure 25C). The ratio of B to T cells in peripheral organs in HBP1 deficient mice is also unchanged (figure 25C). To assess the role of HBP1 in myeloid development bone marrow cells were stained with antibodies against Gr1 and Mac1 (figure 25E). No major differences in the percentages of these cells were detected. These results show that DNA binding and cellular distribution of HBP1 have no major function in development of the haematological compartment of the mice.

Figure 25 Flow cytometric analysis of wildtype, HBP1$^{+/\Delta HMG}$ and HBP1$^{\Delta HMG/\Delta HMG}$ mice

A) Thymocytes were stained with CD4 and CD8 antibodies. Percentage of cells within SP (CD4+/CD8- or CD4-/CD8+), DP (CD4+/CD8+) and DN populations (CD4-/CD8-) are shown.

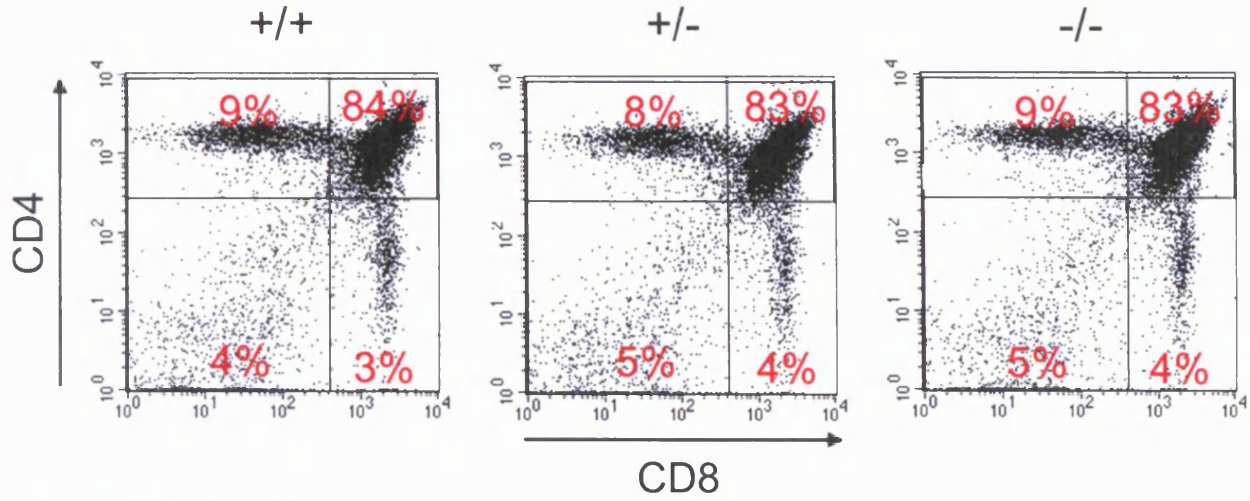
B) Single cell suspensions from lymph nodes were stained with CD4 and CD8 antibodies to allow differentiation of the T cell subpopulations. Percentages are shown.

C) Single cell suspensions from lymph nodes were stained with Thy.1 and CD19 antibodies to allow differentiation of the T cell (Thy.1+/CD19-) and B cell populations (Thy.1-/CD19+). Percentages are shown.

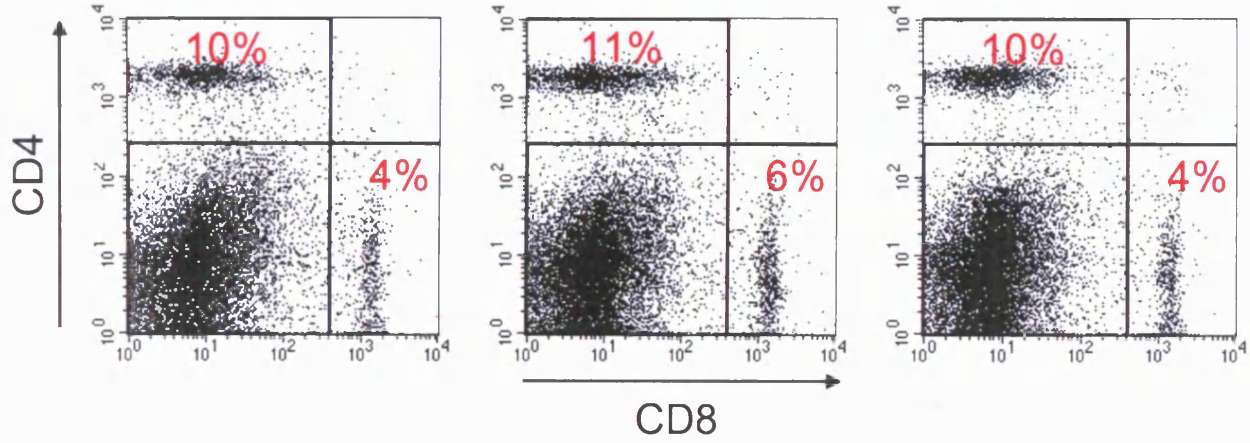
D) Bone marrow cells were stained with CD19 and IgD antibodies. Percentage of cells within mature recirculating CD19+/IgD+ B cells and CD19+/IgD- immature B cells are shown.

E) Bone marrow derived cells were stained with Gr-1 and Mac-1 antibodies. Percentages of cells within gated populations are shown.

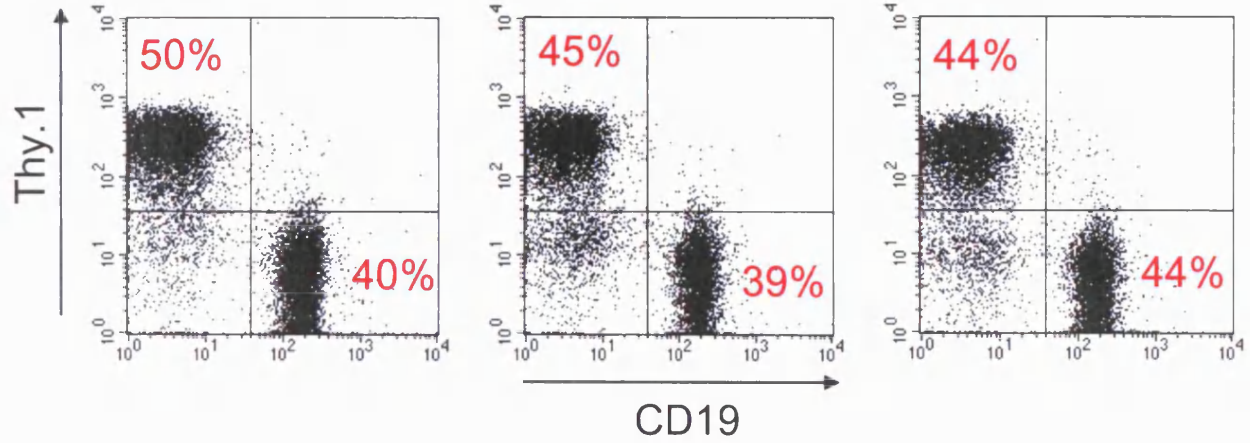
A) Thymus



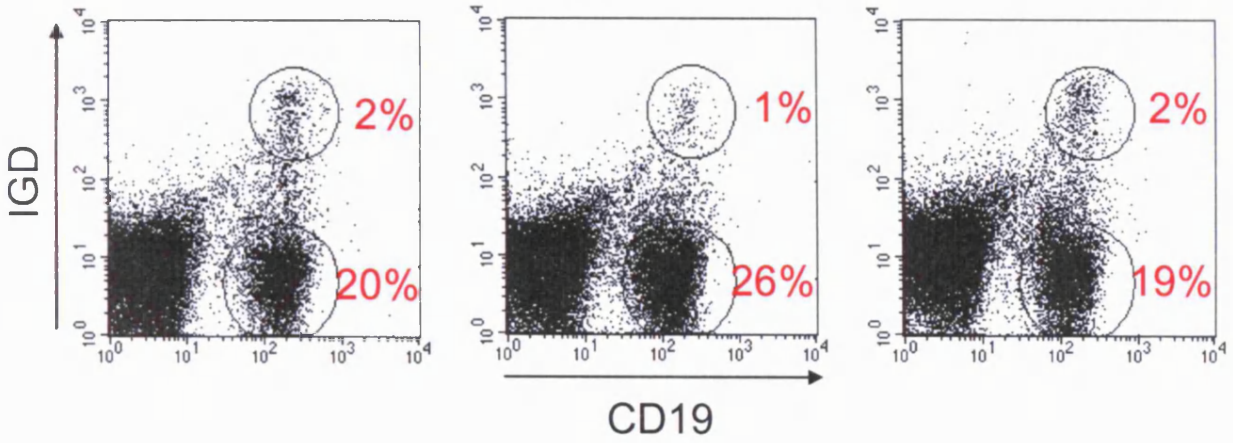
B) Lymph nodes



C) Lymph nodes



D) Bone marrow



E) Bone marrow

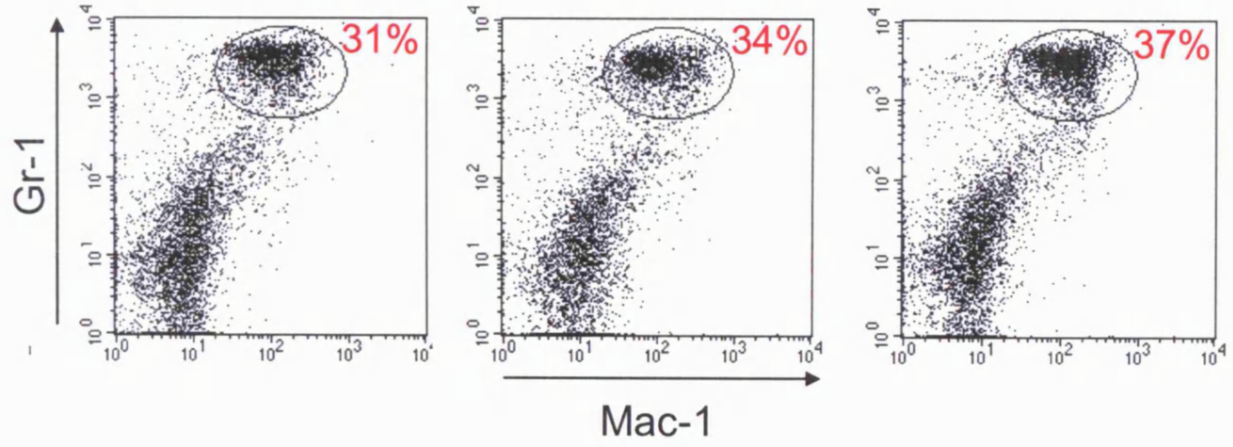


Table 4: Summary of cellularity of HBP1$^{\Delta HMG/\Delta HMG}$ and wildtype mice.

The table shows absolute numbers cells of the different cell types of the organs at 4, 6 and 12 weeks after birth. The mean of triplicates ± one standard deviation is shown.

	4 weeks		6 weeks		12 weeks	
	HBP1$^{\Delta HMG/\Delta HMG}$ (x 1,000,000)	Wildtype (x 1,000,000)	HBP1$^{\Delta HMG/\Delta HMG}$ (x 1,000,000)	Wildtype (x 1,000,000)	HBP1$^{\Delta HMG/\Delta HMG}$ (x 1,000,000)	Wildtype (x 1,000,000)
Lymph node*						
Peripheral T cells	2.2±0.4	2.5±0.8	2.0±0.7	2.6±1.1	1.7±0.4	1.9±0.9
Peripheral CD4 T cells	1.5±0.3	1.7±0.5	1.3±0.4	1.7±0.7	1.1±0.2	1.3±0.6
Peripheral CD8 T cells	0.73±0.1	0.83±0.3	0.67±0.2	0.87±0.4	0.57±0.1	0.63±0.3
Peripheral B cells	1.9±0.4	2.2±0.7	1.8±0.6	2.4±0.9	1.5±0.3	1.7±0.8
Spleen						
Peripheral T cells	15±0.6	20±1.5	17±1.8	15±0.6	15±0.9	15±0.2
Peripheral CD4 T cells	11±0.5	15±1.1	13±1.2	12±0.5	11±0.7	12±0.1
Peripheral CD8 T cells	3.7±0.2	5.0±0.4	3.9±0.4	3.6±1.3	3.4±0.2	3.6±0.05
Peripheral B cells	19±0.8	25±1.9	18±1.9	17±0.6	18±1.1	19±0.2
Thymus						
Double Negative	9.1±0.4	8.9±0.7	4.9±0.4	5.4±0.2	4.7±0.4	5.6±0.2
Double Positive	190±7.6	190±14	140±11	150±4.7	130±9.5	160±4.9
Single CD4 Pos.	20±0.8	20±1.5	16±1.3	18±0.6	16±1.1	19±0.6
Single CD8 Pos.	6.8±0.3	6.7±0.5	4.9±0.4	5.4±0.2	4.7±0.4	5.6±1.7
Bone marrow						
Maturing B cells	2.0±0.8	2.5±0.5	3.2±1.2	2.8±0.4	2.3±1.3	2.9±0.8
Recirculating B cells	0.20±0.09	0.26±0.03	0.27±0.1	0.23±0.04	0.18±0.1	2.3±0.3
Granulocytes and Monocytes	2.5±1.0	3.1±0.7	4.0±1.5	3.5±0.4	2.4±1.3	3.0±0.4

* Cells were taken from Mesenteric lymph nodes

** Cells were taken from the bones of one back leg

Effect of HBP1 expression on the apoptosis of thymocytes

The transfection studies described earlier in this thesis suggested that HBP1 over expression might induce apoptosis. To study the potential role of HBP1 in apoptosis we measured the response of T cells from HBP1 wild type and targeted mice to apoptotic stimuli. Apoptotic cell death can occur via several distinct pathways, which are regulated by different molecules. In this study, the following apoptotic stimuli were used: γ-irradiation, which causes DNA damage, and dexamethasone, which activates glucocorticoid receptors leading to cell death in thymocytes. Thymocytes from $HBP1^{+/+}$, $HBP1^{+/\Delta HMG}$ and $HBP1^{\Delta HMG/\Delta HMG}$ were treated with apoptotic stimuli, the samples harvested at various time points and the amount of apoptosis determined by propidium iodide staining of DNA followed by flow cytometry analysis (Nicoletti et al., 1991). It has to be noted that the experiment was done only once with each sample in triplicate. γ-radiation induces p53 dependent apoptosis through DNA damage, and thymocytes in particular are highly sensitive to this form of damage. The level of apoptosis in the thymocytes of the HBP1 targeted mice was determined following treatment with 5Gy of γ-radiation (figure 26A). No difference in the rate of cell death was seen in thymocytes of $HBP1^{+/+}$ compared to $HBP1^{\Delta HMG/\Delta HMG}$. This indicates that absence of the HMG box from HBP1 does not affect the rate of apoptosis, which is induced through γ-radiation. Thymocytes from the HBP1 targeted mice were also exposed to dexamethasone, which induces p53-independent apoptosis. Again, no difference in the rate of cell death was seen in thymocytes of $HBP1^{+/+}$ compared to $HBP1^{\Delta HMG/\Delta HMG}$ (figure 26B). This indicates that the absence of the

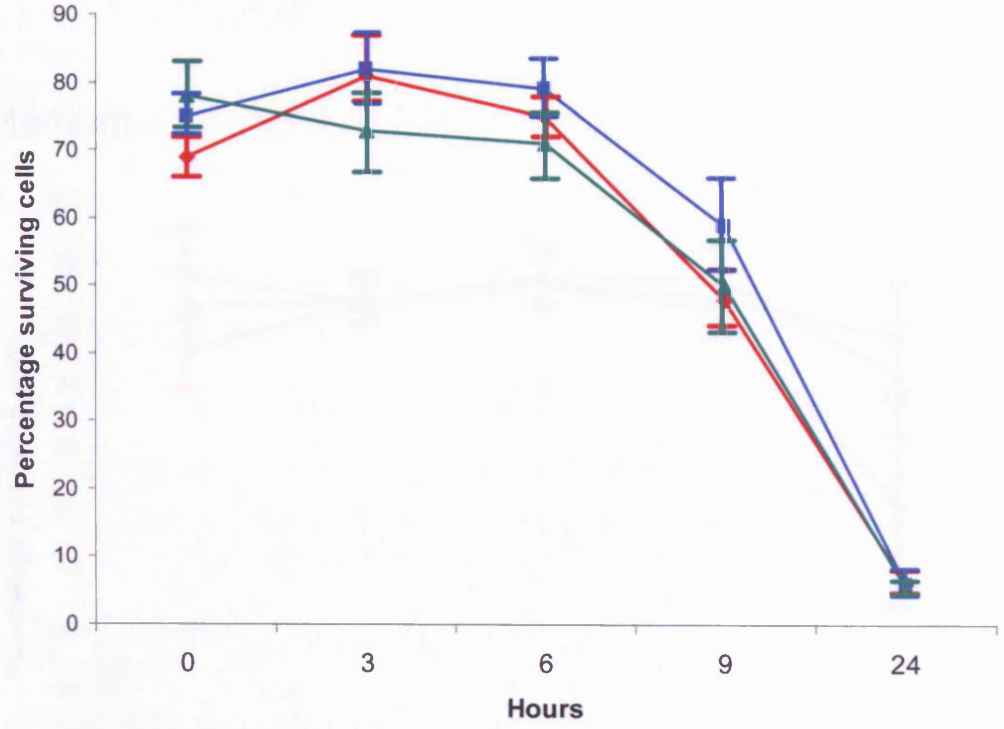
HBP1 HMG box binding activity does not alter the rate of apoptosis, which is induced through dexamethasone. As a control, thymocytes from the HBP1 targeted mice were placed in culture without any apoptotic stimuli. The rate of survival of the thymocytes from all mice was equal (figure 26C).

These results indicate that the thymocytes from HBP1$^{\Delta HMG/\Delta HMG}$ mice are as sensitive/insensitive to apoptotic signals as wild type thymocytes. It was concluded that at least the DNA binding activity of HBP1 has no function in the activation of or protection against apoptosis.

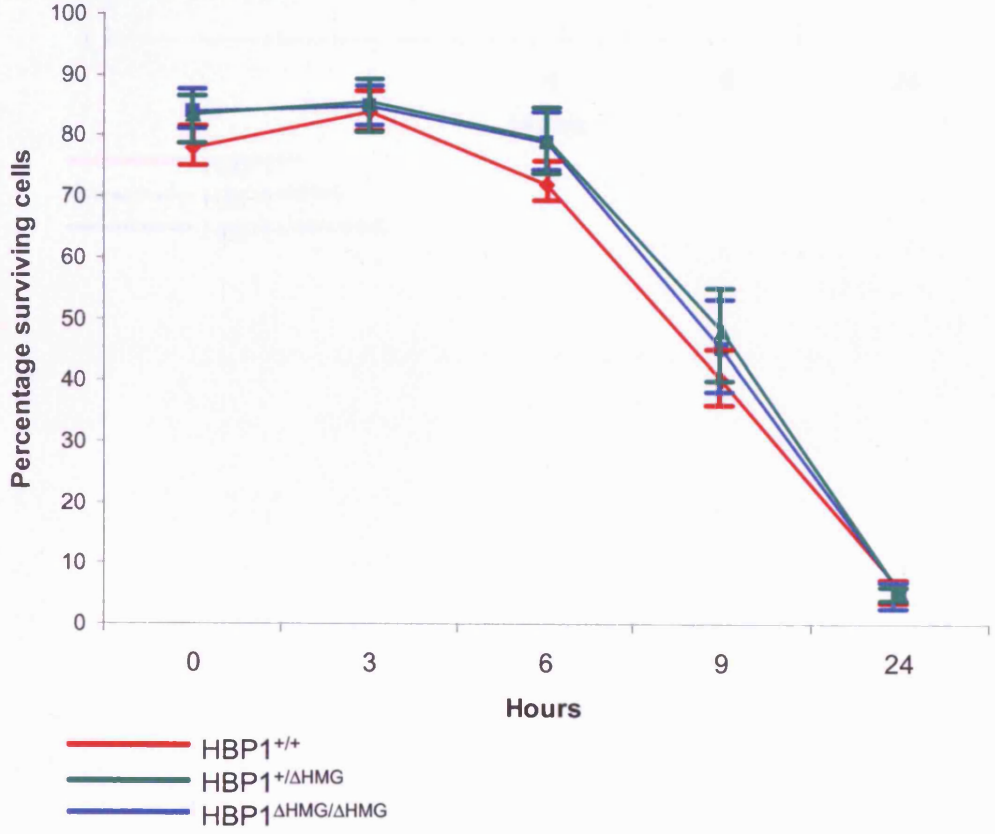
Figure 26 HBP1$^{\Delta HMG/\Delta HMG}$ thymocytes show wildtype levels of apoptosis

Total thymocytes from HBP1$^{+/+}$ (red line), HBP1$^{+/\Delta HMG}$ (green line) and HBP1$^{\Delta HMG/\Delta HMG}$ (blue line) were cultured in vitro for up to 24 hours. Charts show the percentage of surviving cells per time point. Studies were carried out following treatment with either (A) 5 Gy γ-irradiation, (B) 5μM Dexamethasone or (C) untreated controls. Error bars indicate mean ± SD.

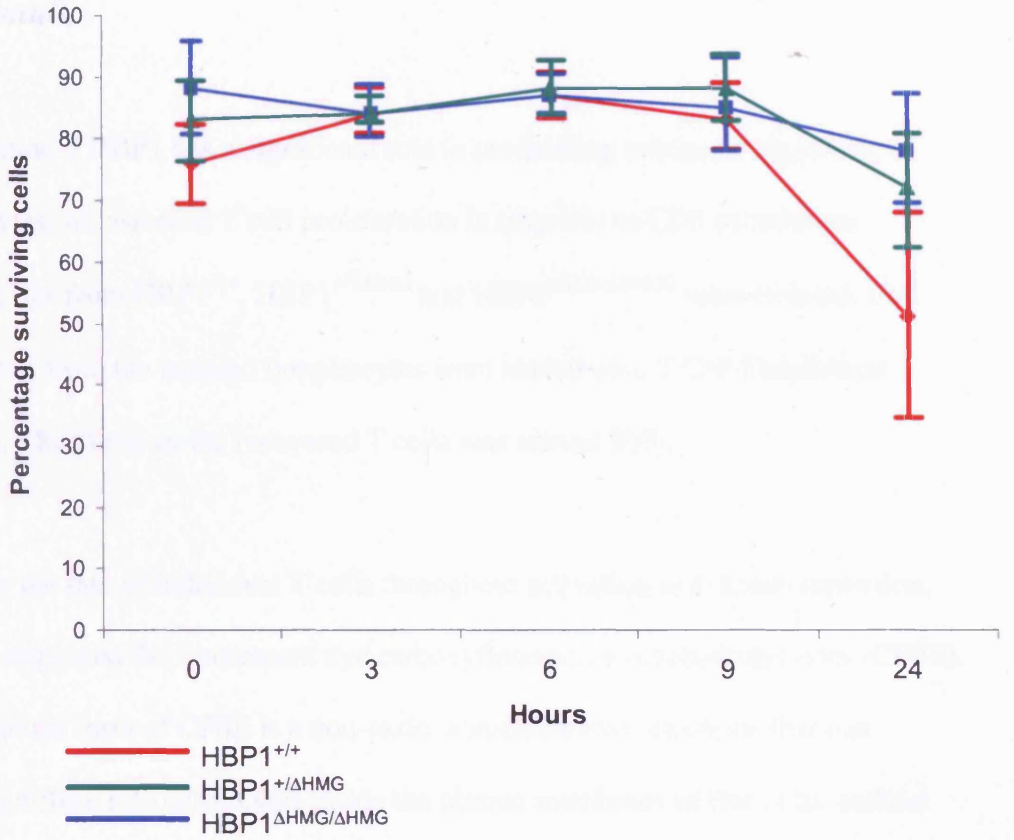
A: Radiation



B: Dexamethasone



C: Medium


100
90
80
70
60
50
40
30
20
10
0
Percentage surviving cells
0
3
6
9
24
Hours
HBP1+/+
HBP1+/ΔHMG
HBP1ΔHMG/ΔHMG

HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes exhibit enhanced proliferative response to TCR stimuli.

To determine if HBP1 has a significant role in modulating mitogenic signalling of lymphocytes, we assessed T cell proliferation in response to CD3 stimulation. Lymphocytes from HBP1$^{+/+}$, HBP1$^{+/\Delta HMG}$ and HBP1$^{\Delta HMG/\Delta HMG}$ were isolated. Cell suspensions from the isolated lymphocytes were loaded onto T Cell Enrichment Columns. The purity of the recovered T cells was around 90%.

To follow the fate of individual T cells throughout activation and clonal expansion, we have employed the fluorescent dye carboxyfluorescein succinimidyl ester (CFSE). The di-acetate form of CFSE is a non-toxic, non-fluorescent molecule that can passively diffuse into cells. Once inside the plasma membrane of live cells, cellular esterases cleave acetyl groups from the di-acetate form of CFSE to create the active fluorophore (CFSE) that absorbs and emits light at wavelengths characteristic to its fluorescein moiety. The highly amine-reactive succinimidyl ester then forms dye-protein adducts which are retained within the cells or at the cell surface. The result is the very bright labelling of cell populations that is stable over a span of weeks. This labelling is universal within the T cell compartment, as all CD4 or CD8 positive cells fluoresce brightly before proliferation (data not shown). CFSE segregates equally between daughter cells upon cell division, resulting in the sequential halving of cellular fluorescence intensity with each successive generation. When analyzed by flow cytometry, this sequential halving of fluorescence is visualized as reduced

fluorescent intensity of populations of cells, and can be used to track cell division in populations of proliferating cells.

Figure 27A shows a histogram of CFSE intensity from a representative experiment. Three days post-stimulation both CD4 and CD8 HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes exhibited a greater proliferation, as shown by increased CFSE dilution, compared to HBP1$^{+/+}$ lymphocytes (figure 27A). This affect was more pronounced in the CD8 cells compared to the CD4 cells. The CFSE profile were used characterize two parameters; the percentage of cells that had been triggered to divide (percent divided) and the mean number of divisions among those cells that had divided at least once (proliferation index or burst size).

The percent divided, that is the percentage of the cells of the original sample that have divided, is shown in figure 27B. No differences between the HBP1$^{\Delta HMG/\Delta HMG}$ compared to wildtype were measured for the first 48 hours. However, after 72 hours of culture, a clear difference was measured. With greater proportion of CD4 and CD8 cells from HBP1$^{\Delta HMG/\Delta HMG}$ being triggered into division compared to wildtype cells. The difference seen was more significant for CD8 T cells compared to CD4 T cells. Thus, more lymphocytes from HBP1$^{\Delta HMG/\Delta HMG}$ mice are triggered into division compared to wildtype mice.

The proliferation index, or the average number of divisions that a dividing cell has undergone, was equal for CD4 HBP1$^{\Delta HMG/\Delta HMG}$ compared to wildtype CD4 lymphocytes and may be slightly increased for CD8 HBP1$^{\Delta HMG/\Delta HMG}$ compared to

wildtype CD8 lymphocytes (figure 27C). Thus, once a cell is triggered to go into the cell cycle they undergo about the same amount of divisions after 72 hours.

Taken the data together this indicates that both CD4 and CD8 HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes are more responsive to TCR signalling but once the cells are cycling, the cells undergo a similar amount of division (although this may be slightly increased for the CD8 HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes). This might indicate that HBP1 regulates the cell cycle entry of resting cells but once they are in cycle, other control mechanisms take over.

A plausible explanation why HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes respond with increased cell proliferation might be that HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes are in a "pre-activated" state and thus predisposed to immediate response. A few gene-targeting studies have described cases of pre-activated lymphocytes that have elevated CD69 and CD25 levels (Grossman et al., 1999; Kuo et al., 1997). To explore this possibility, FACS analysis of HBP1$^{+/+}$ and HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes was first performed for CD69, which is highly expressed on activated T cells. CD69 was absent on wild-type cells which did not get stimulated by CD3. Intermediate levels of CD69 expression was measured on non-stimulated HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes (figure 28). This indicates that these cells might be pre-activated. 24 h post-stimulation CD69 was up regulated in both HBP1$^{\Delta HMG/\Delta HMG}$ and wild type cells but the level of up-regulation was greater in HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes. Similar results were seen for the high affinity IL2 receptor (CD25). Non-stimulated knockout lymphocytes show intermediate levels of CD25 while expression was absent on wild-type cells. 48 hours post activation CD25 was up regulated in both HBP1$^{\Delta HMG/\Delta HMG}$ and wild type cells but again the increase of

CD25 expression was greater in the knockout cells compared to wild-type cells (figure 28). Taken the results together this might indicate that HBP1$^{\Delta HMG/\Delta HMG}$ cells are more responsive to signalling through the TCR and this might be due to a pre-activated phenotype.

Figure 27 T cells from HBP1$^{\Delta HMG/\Delta HMG}$ are poised for activation

A) CFSE profile after 72 hours

Lymphocytes from HBP1$^{\Delta HMG/\Delta HMG}$ mice (single black line) and wild-type mice (filled grey area) were stained with CFSE prior to activation. Non-stimulated lymphocytes do not divide whereas anti CD3 activated HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes exhibit enhanced entry into the cell cycle as shown by a greater decrease in CFSE.

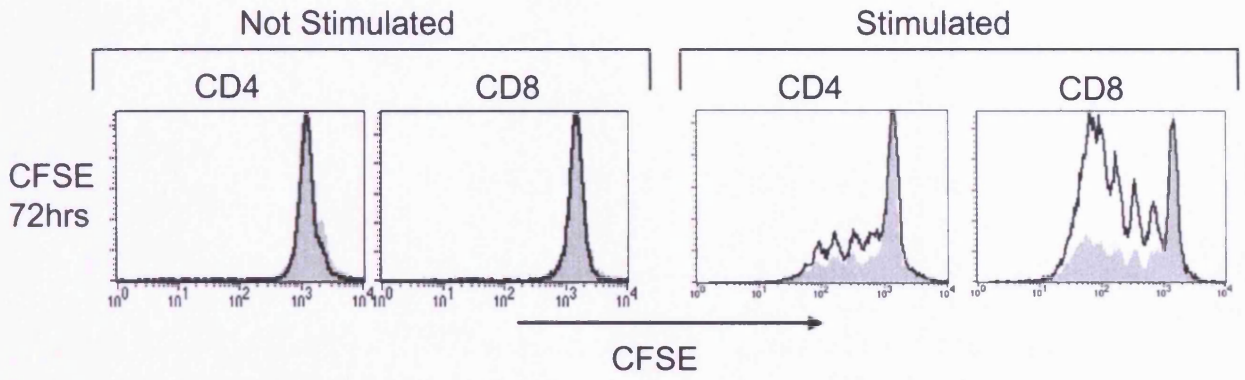
B) Percent divided after stimulation

More cells underwent division for both CD4 and CD8 HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes compared to wildtype as shown by increase in the % divided. Error bars indicate mean ± SD.

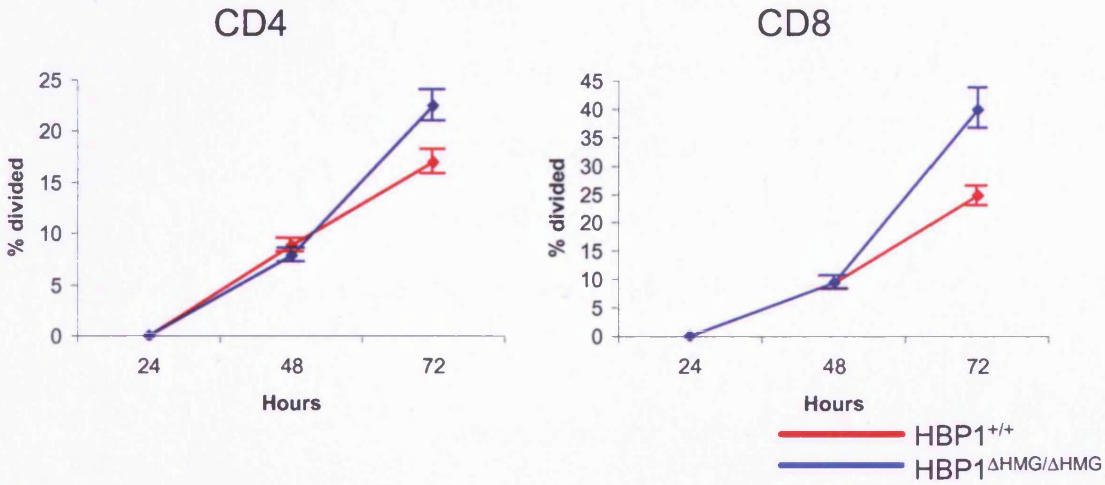
C) Proliferation index after 72 hours

The average number of divisions that the cells, which divided, underwent is similar for the CD4 HBP1$^{\Delta HMG/\Delta HMG}$ and wildtype lymphocytes. The CD8 HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes have a slight increased proliferation index compared to wildtype lymphocytes. Error bars indicate mean ± SD.

A) CFSE profile after 72 hours.



B) Percent divided after stimulation



C) Proliferation Index after 72 hours

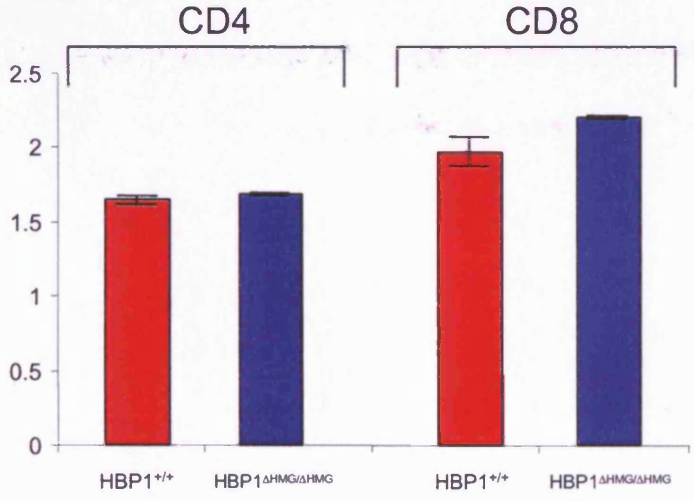
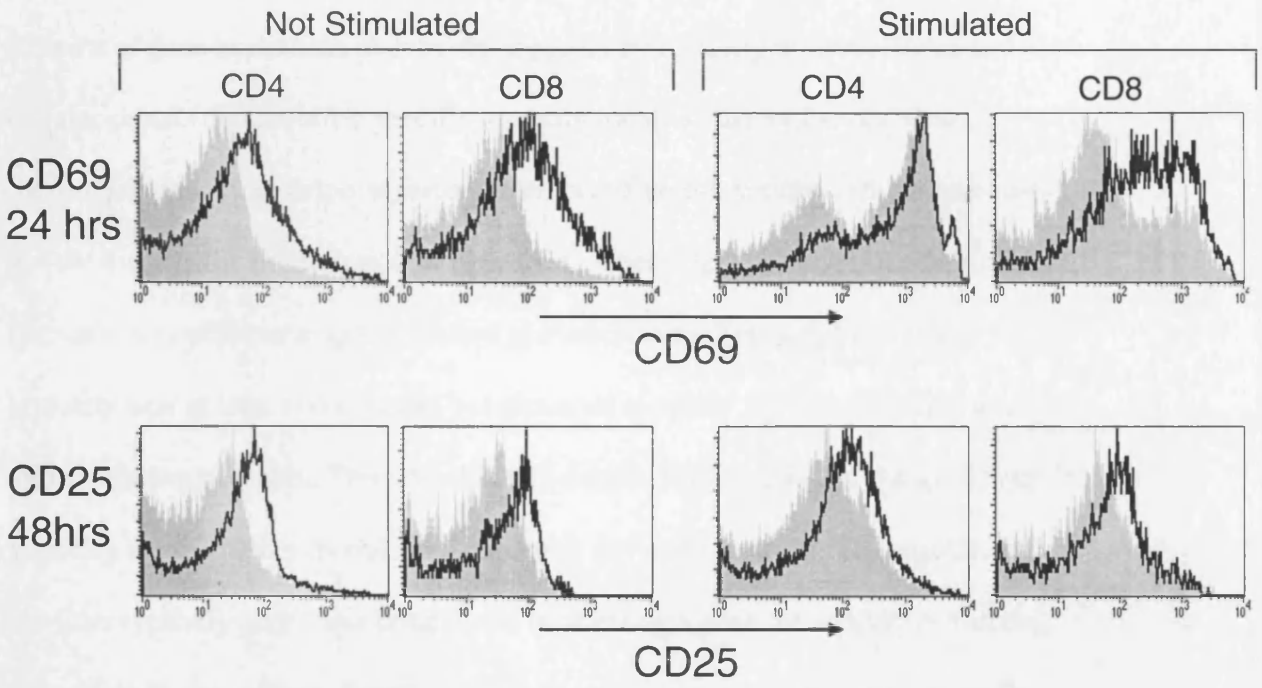


Figure 28: HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes have up regulated levels of CD69 and CD25

Lymphocytes from HBP1$^{\Delta HMG/\Delta HMG}$ (single black line) mice exhibit up regulated levels of CD69 and CD25 prior and after activation with ant-CD3 compared to wildtype littermates (filled grey area). After stimulation for 24 hours for CD69 or 48 for CD25 HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes still exhibit higher levels of CD69 and CD25 compared to wildtype.


Not Stimulated
Stimulated
CD4
CD8
CD4
CD8
CD69
24 hrs
CD69
CD25
48hrs
CD25

Discussion

Domains of HBP1

Patterns of gene expression in cells are triggered by a variety of intracellular and extra cellular signals that mobilize specific transcriptional regulatory factors. Once marshalled, these transcription factors interact and recruit various cellular machines to govern directly the transcription of the specified genes, prescribing both the time and the scale of positive or negative control of transcription. Transcription factors typically bear at least two essential yet separable modules: the DNA-binding domain and a regulatory domain. The DNA binding domain imparts most of the specificity in targeting the regulatory module to a particular site in the genome. The regulatory modules typically play a less critical role in selecting a gene for regulation; instead, they mediate their effects directly or through interaction with other proteins on the gene to which they are delivered. The DNA-binding modules of transcriptional regulators consist of a wide variety of structural folds and many have been well characterized from both a structural and functional standpoint. By contrast, regulatory modules are less well defined and only a few have been characterized structurally. As described in the introduction of this thesis the transcription factor HBP1 contains two modules that are recognized by BLAST analysis, the AXH module and the HMG box. The AXH module is a novel regulatory element identified in HBP1 and in SCA1. The gained insights in the function and properties of this domain in terms of its structural

properties and stability will be discussed in the following section. This is followed by a functional characterisation of the HMG box, the DNA binding domain of HBP1, and its function in sub-cellular localisation of the protein.

The AXH module of HBP1

Expression of the AXH domain yields an independently folded polypeptide with the N and C termini juxtaposed in space. It was shown that deletion mutants that cut out even a few residues from the N and C termini adjacent to the AXH domain led to unfolded proteins. Thus, the stability of the AXH module is affected by the presence of interactions with adjacent poly-peptides. This implies that boundary selection is clearly important. Incorrect selection of the boundary of AXH by a couple of residues removes several interactions that contribute to the overall stability of the structure. This has local and global effects on the stability and alters related biophysical characteristics such as the unfolding rates.

This observation has direct implications for interpreting mapping studies of the interaction sites of HBP1 with other proteins. Only constructs that comprise the full-length AXH module are able to retain binding, whereas truncation of AXH abolishes binding as expected, if its tertiary structure is disrupted.

The AXH domain of HBP1 is monomeric. This is in stark contrast with the AXH domain of SCA1, which forms dimers in solution. A dimerization domain of ATX1 falls within a region spanning residues 495–605 (Burright et al., 1997) and this region

partially overlaps with the AXH domain of ATX1 (568-689). Thus, it is likely that the residues responsible for the dimerization of AXH of SCA1 fall within a 37 aa region (568-605). The dimerization is not an intrinsic feature of AXH, since the AXH of HBP1 is monomeric. It is conceivable that the non-conserved residues within the identified 37 aa dimerization domain are the residues involved in protein dimerization. Therefore, it is unlikely that the AXH domain of HBP1 interacts with the AXH domain of AXT1.

In conclusion, we have used the sequence similarity between regions of two otherwise unrelated proteins to obtain information about their domain architecture. The AXH motif is a novel protein–protein binding domain. The high degree of similarity between the domains of ATX1 and HBP1 may indicate that both proteins bind to similar targets. One of the targets the has been identified to bind to both ATX1 (Fernandez-Funez et al., 2000) and HBP1(Swanson et al., 2004) is the Sin3 complex. Deacetylation of histones by the SIN3 complex is a major mechanism utilized in eukaryotic organisms to repress transcription. Thus, the interaction between Sin3 with AXH from both ATX1 and HBP1 implicates both genes in transcriptional repression.

Sub-cellular localisation of HBP1: The HMG box and identification of NLS

The DNA binding domain is necessary for the correct targeting of the transcription factor to a particular site in the genome. In eukaryotic cells, the nuclear envelope generates two distinct cellular compartments that separate transcription and genome (nuclear) from protein biosynthesis (cytoplasm). Thus, a given transcription factor has

to be transported into the nucleus to ensure its function. Before we set out to define the region necessary for the nuclear transport, we confirmed that transcription factor HBP1 is a nuclear protein. To visualize the protein it was fused to the fluorescent protein EYFP that allows the protein to be detected by fluorescence microscopy. The fluorescence of the fusion protein was only detected in the nucleus. This confirmed that HBP1 is a nuclear protein.

The nuclear membrane of eukaryotic cells is freely permeable to solutes of up to about nine nm (e.g., 40- 60-kDa proteins). Transport of larger molecules through nuclear pores is signal-mediated, involves shuttle molecules, and requires energy. These larger molecules contain peptides composed of karyophilic clusters of arginines and lysines, which signal transcription factors to specialized transporter molecules in the pore complex of the cytoplasm. These karyophilic clusters are called nuclear localization signals (NLS). Two types of NLSs have been described in the literature, one contains four arginines and lysines within a hexa-peptide and the other comprise two clusters of basic amino acids separated by a non-basic peptide. Analyses of the sequence of the HBP1 gene lead to the identification of a bipartite NLS motif, consisting of two clusters of basic amino acids separated by a non-basic peptide, within the HMG box (435-KRPMNAFMLFAKKYRVE-451). These sequences are highly homologous with corresponding NLS motif in other HMG domain proteins. To confirm that the identified NLS was active as an NLS two mutants of HBP1 were fused to EYFP. One mutant consists of just the HMG box while the other mutant contained the rest of the protein (figure 7). Deletion of the HMG box resulted in accumulation in the cytoplasm. However, deletion of the HMG box did not prevent the fusion protein from entering the nucleus. This could be due to the size of the

fusion protein that is around 60 kDa, small enough to enter the nucleus by diffusion through the nuclear pore. These results demonstrate that the HMG box is a critical determinant for nuclear localization.

Thus, the HMG box provides the DNA binding domain and the necessary signals for the transport of HBP1 from the cytoplasm to the nucleus. The co-localization of the DNA binding domain and the NLS to one part of the gene is not uncommon. A survey showed that the NLS of ~70% of nucleic acid-binding proteins are coincident with their nucleic-acid binding domain. It is not known why nuclear import and DNA binding are so closely localized on these molecules, but it has been speculated to be an evolutionary consequence of exon shuffling.

The location of a transcription factor within the nuclear space can be important for its function. The HMG box, the DNA binding domain of HBP1, is necessary for the targeting of HBP1 to particular sites in the genome. When the EYFP-HMG box was expressed, it was localized as fluorescent speckles in the nucleus. However, when the full length EYFP-HBP1 was expressed the protein seemed to show a different localization pattern. Unfortunately, due to the lack of antibodies against HBP1 we could not perform immuno-fluorescence staining on endogenous expressed HBP1. Therefore, we could not exclude that the observed fluorescent speckles are an artefact due to the formation of inclusion body like structures upon over expression of the EYFP fusion proteins. If the observed phenotype is not an artefact, the significance of this might be that the AXH, or other protein binding sites in the HBP1 gene, can form complexes with other proteins that affects the function of the gene and the spatial localization of HBP1 in the nucleus.

Besides the difference in localization, there was also a functional difference between the full-length fusion protein and the two mutant fusion proteins. 72 hours post transfection no fluorescence was detected in the cells transfected with the full-length fusion protein. Furthermore, time-lapse experiments showed that the transfected cells rounded up and detached from the surrounding cells. This might indicate that over expression of the EYFP-HBP1 causes apoptosis or is at least toxic for the cells. At this moment in time, we do not know if this phenotype is due to HBP1 or EYFP. It has been reported that fluorescent fusion proteins can be harmful or even toxic to cells in transient transfection experiments. Although, when EYFP-HMG or even EYFP-ΔHMG was expressed high levels of fluorescence were detected in the cells after 72 hours. This indicates that these mutants are not harmful for the cell.

From these experiments, we conclude that the HMG box provides the signals for the nuclear localisation and is necessary to target HBP1 to certain sites in the genome, while the rest of HBP1 mediates its effects directly or through interaction with other regulatory proteins. Furthermore, when both domains are uncoupled the resulting two proteins (HMG and ΔHMG) are not harmful to the cell.

Gene targeting

Conditional gene targeting is increasingly used to create loss-of-function mouse models, and a standard approach involves the Cre-loxP recombination system. This approach involves creating a floxed conditional allele, designed to remain functionally equivalent to the wildtype, and then inactivating it in vivo in a cell type-specific and

temporally controlled fashion, using transgenes expressing Cre recombinase. First, the generation of two Cre recombinase mouse lines will be discussed. This is followed by a discussion of mice in which an essential part of the HBP1 is floxed.

Cre recombinase transgenic mice

The generation of conditional mutants is becoming increasingly important in determining the physiological role of specific gene products. To date, the Cre/loxP based approach has been the most successful method to generate defined mutations in a spatially and temporally controlled manner. Here two independent Cre/loxP based systems are described for the generation of conditional mutants within the haematopoietic system. The first places Cre expression under the control of the Vav1 regulatory elements which drive expression in all haematopoietic lineages (Ogilvy et al., 1999), presumably including the haematopoietic stem cells, thus allowing pan-haematopoietic targeting. In the second system Cre expression is controlled by the hCD2 LCR and allows targeting to be restricted to lymphocytes only (Lang et al., 1988).

Cre expression in transgenic mice

Cre activity in both the Vav-iCre and the hCD2-iCre transgenic mice was demonstrated by the subsequent mating of the established transgenic lines with the ROSA26-EYFP transgenic mouse line (Srinivas et al., 2001) creating double transgenic offspring. The ROSA26-EYFP mouse line carries the EYFP gene under the

control of the ubiquitous and constitutive expressed ROSA26 promoter. Inserted amid the ROSA26 promoter and the EYFP gene are the loxP sequences flanking a strong transcriptional stop sequence to prevent transcriptional read-through. In the presence of the Cre-recombinase, the stop sequence is removed and the EYFP gene is expressed. Thus, fluorescence microscopy reveals if any Cre-mediated recombination events have occurred. The finding of EYFP expression demonstrates that the reporter gene chosen is functional. Measuring fluorescence is not without its pitfalls and one has to be careful to avoid false positive and false negative results.

One should generally show caution when measuring fluorescence because of auto fluorescence that can be interpreted as a positive signal. The included single transgenic ROSA26-EYFP mice therefore served as control that the fluorescence observed was indeed caused by EYFP expression and not background fluorescence.

Strebel et al. (Strebel et al., 2001) showed that GFP transfected haematopoietic cell lines undergoing apoptosis or necrosis have a decrease in fluorescence. Decreased EGFP fluorescence appears to be an early and very sensitive marker of cell death. This could lead to false negative cells when all the cells analysed are presumed to be alive. Indeed, when we took lymphocytes from our fluorescent mice expressing either GFP or EYFP and left these cells for couple of hours at room temperature to die, we detected, by flow cytometry, a significant proportion of GFP or EYFP negative cells, which were not present in the fresh sample (data not shown). T cells were more prone to loss of fluorescent signal compared to B cells and this may reflect the rate of death for these cells in vitro. Steff et al. (Steff et al., 2001) showed that the amount of fluorescent protein was similar in living and dying cells ruling out the possibility that

fluorescent proteins can leak out of cells or are degraded. They suggested that fluorescent proteins are quenched by modifications in the intra-cellular milieu. Thus, dying cells expressing fluorescent proteins lose their fluorescence and can cause false negative results when not excluded from the analysis.

The results obtained by mating the iCre transgenic mice lines with the ROSA26-EYFP mouse line showed expression of EYFP in all lymphoid tissues. Except for expression of EYFP in the testis and ovaries for the Vav-iCre/EYFP mice and variegated expression in the testis for the hCD2-iCre mice, no EYFP expression was detected in other tissues.

The expression of EYFP in the testis of the iCre transgenic lines was unexpected. The Vav-iCre transgenic mice used an expression vector that previously had been reported to be inactive in the testis. The testis from hCD2-iCre/EYFP mice displayed a variegated appearance indicating that Cre had been expressed in a proportion of the cells in this organ. It should be noted here that expression of hCD2 has never been noted in the testis of transgenic mice, nor has expression of GFP been detected in the testis of hCD2-GFP mice (de Boer et al., 2003).

It is not likely that this aberrant expression in testis is due to the site of integration of the transgene, as it is observed in two independent transgenic lines represented here, as well as in the other Vav-iCre lines that we have generated.

Prokaryotic sequences can have erratic performance in transgenic animals (Cui et al., 1994). Those sequences contain numerous potential binding sites for mammalian

transcription factors. These complexes formed between the prokaryotic DNA and the mammalian transcription factors could interfere with proper assembly on linked mammalian regulatory elements. Thus, it is likely that iCre contains DNA sequences that are able to activate transcription of the Vav-iCre and hCD2-iCre transgenes in the testis of these mice. Besides inappropriate expression in the testis, a number of other Vav-iCre lines that we generated exhibited mosaic expression of Cre in several non-haematopoietic tissues, including gut, brain, and muscle (data not shown). These results support the notion that some prokaryotic transgenes are able to cause aberrant transcriptional regulation of transgenes and neighbouring genes at the site of integration.

Besides inappropriate activation of loci prokaryotic sequences contain CG sequences, which are potential mammalian methylation sites, which can contribute to the creation of a silent state (Clark et al., 1997). The inactivation might be due to stable repressed complexes, hetero-chromatinisation, and/or methylation. Support that Cre can induce transcriptional silencing became evident in a number of hCD2-iCre transgenic lines that we have produced. The hCD2 LCR is able to direct tissue specific expression of mammalian transgenes in a position independent copy number dependent manner. However, in a number of hCD2-iCre lines, expression of Cre was either mosaic or not detectable (data not shown). In these mice, Cre was able to overcome the powerful regulatory elements of the hCD2 LCR and induce variegation, possibly by enforcing hetero-chromatinisation of the transgenic locus at the site of integration.
A transgenic line expressing Cre under the control of the hCD2 regulatory elements has been previously described, and according to expectation Cre activity was detected only in lymphoid organs (Wilson et al., 2002). However, the detection system used in

this study did not allow the assessment of Cre expression in individual cells and therefore, the issue of variegation was not addressed.

In addition to the transcriptional deregulation, we also noted toxic effects of Cre expression. The Vav-iCre and hCD2-iCre transgenic lines presented in this report showed no detectable differences in thymic cellularity or development when compared to non-transgenic littermate controls. However, in a number of other hCD2-iCre lines and one Vav-iCre line the size of the thymus was decreased approximately ten fold when compared to non-transgenic littermates (unpublished data). High levels of Cre expression have been reported to be toxic in eukaryotic cells, possibly due to chromosomal rearrangements caused by recombination between cryptic 'pseudo-loxP' sites naturally found within the genome (Thyagarajan et al., 2000). Alternatively, integration of the transgene into the genome of these mice may have disrupted a gene important for thymic development. However, this possibility seems unlikely as the phenomenon was noted in a number of independently generated lines.

Pattern of Cre expression in T cells

EYFP expression in T cells from the R26R-EYFP and iCre double transgenic mice was analysed by flow cytometry. This showed that EYFP was expressed in all peripheral T cells of both Vav-iCre and hCD2-iCre double transgenic mice. T cell development is normally analysed by the expression of CD4 and CD8 on the thymocyte surface. When thymocytes were stained with these markers all phases from double negative to single positive showed close to hundred percent expression of

EYFP in both Vav-iCre and hCD2-iCre double transgenic mice. Even though great care was taken to avoid dying cells from the analysis, a couple of cells were detected that have lost their fluorescence. These negative cells are most likely dying cells that are not longer fluorescent (Steff et al., 2001; Strebel et al., 2001).

It is possible to analyse further the double negative population by staining these cells with CD44 and CD25. This allows us to follow the development of the earliest T cells organised from DN1 to DN4. In both iCre transgenic lines, even the earliest progenitors to be found in the thymus express the EYFP reporter gene. However, analysis of hCD2-iCre mice revealed a small population of EYFP negative cells within the DN1, DN2 and DN3 gates. In hCD2-GFP mice, the amount of GFP negative cells decreased from the DN1 stage until the DN4 stage were all cells were GFP positive. Less GFP negative cells were found in DN1 from the hCD2-iCre R262R EYFP mice compared to hCD2-GFP mice. These differences could be due to a number of factors. The hCD2-iCre transgenic locus might be integrated in a more open chromosomal locus compared to the hCD2-GFP locus. This could allow earlier expression from the hCD2-iCre because it is more accessible to transcription factors. The differences could be in the mechanism used to report expression of each construct. Expression of Cre results in the deletion of the R26R-EYFP stop signal, which is a permanent modification that results in the expression of EYFP. All subsequent daughter cells therefore express EYFP and at levels determined by the ROSA26 locus. In contrast, expression of GFP is directly under the control of the hCD2 regulatory elements and will only be expressed when the hCD2-GFP transgene is actively being transcribed and at levels determined by the hCD2 regulatory elements.

Pattern of Cre expression in B cells

All B cells analysed in the VAV-iCre and hCD2-iCre double transgenic mice expressed EYFP. This indicates that the deletion occurs early in B cell development.

Mosaic expression of GFP was seen in hCD2-GFP peripheral B cells, and this could account for the mosaic appearance in fluorescence within the Peyer's patches of these mice. In contrast, all $CD19^{+}/IgD^{-}$ immature B cells expressed GFP. The fact that all immature hCD2-GFP B cells express GFP while the expression mature B cells is variegated suggests that the hCD2-GFP construct is active in early stages of B cell differentiation, but gets turned off during later stages of B cell development.

Pattern of expression in other haematopoietic lineages

It has been reported previously that the Vav1 regulatory elements were able to direct transgene expression in all haematopoietic cell types. Accordingly, expression of the EYFP reporter was noted in all granulocytes, monocytes and erythrocytes indicating that iCre had been expressed in all these cells and possibly their progenitors.

EYFP in erythrocytes showed a more varied profile in the Vav-iCre double transgenic mice. This can be due to the gradual loss of EYFP after the enucleation of these cells. This is because no new proteins are made in these cells and the half-life of

erythrocytes is longer than the half-life of EYFP. In agreement with the known pattern of hCD2 expression in transgenic mice, no expression of EYFP or GFP was detected in these cell types in either hCD2-iCre/EYFP of hCD2-GFP transgenic lines, respectively.

Genomic organization of HBP1

To gain better insights into the function of the HBP1 gene, we determined its genomic structure. These studies should provide valuable information for our gene knockout experiments aiming at dissecting the physiological roles of HBP1, and should facilitate our understanding on the transcriptional regulation of the HBP1 gene.

The initial strategy for obtaining HBP1 genomic locus depended on screening a lambda phage library. Phage clones containing the 5' end of the HBP1 gene could not be obtained from critical phage clones, so a PAC clone, was used to obtain the HBP1 genomic locus. Fragments from the PAC clone, were sub cloned to construct a sequencible contig. The sub clones were sequenced using a modified primer walking protocol. A map of the HBP1 gene was constructed utilizing the known mRNA sequence to establish the exon–intron organization of the gene (figure 18).

The mouse HBP1 gene is approximately 23 kb, coding for 10 exons and 9 introns. Exons 1 to 9 and a small portion of exon 10 code for the 514 amino acid HPB1 protein.

Different domains of the HBP1 protein have mapped to the cDNA sequence. Combining that information with the data about the gene's organization, as shown here, reveals that the first RB binding domain of the protein falls within exon 2. While the second RB binding domain is within exon 6. The AXH domain of the protein is encoded by exons 5 to 7. Finally the HMG box of HBP1has been mapped to exons 8 and 9.

The role, if any, of the remainder of exon 10 remains to be determined. A comparison of human and mouse orthologs of HBP1 revealed that the 3' untranslated region of the HBP1 gene is highly conserved (95% homology) between mouse and human. The significance of highly conserved 3' UTR is not known. The HBP1 contains a 1.1 kb long 3' UTR which can be spliced. This alternative splicing alters the sequence of the five most C-terminal amino acids and includes or omits 1 kb of 3' untranslated region to the HBP1-encoding mRNAs. In general, 3' UTRs act in cis to regulate mRNA translation, stability, and localization through association with regulatory proteins or with antisense RNA. It is interesting that the identified antisense RNA by (Yelin et al., 2003) aligns to the region that is alternatively spliced. It is possible that the 3' untranslated region of HBP1 contains binding sites for regulatory proteins. The antisense RNA may enhance or abolish the establishment of complex protein interactions on the 3' UTR that can affect HBP1 mRNA.

The genomic data provided in this thesis have been used to prepare the targeting construct for the generation of HBP1 knockout mice. This will be discussed in the following section.

Floxing the HMG box of HBP1

To inactivate the HBP1 gene in a temporal and positional manner an essential part of the gene one or several exons in general need to be flanked with loxP sites by homologous recombination. It is, however, necessary to ensure that the loxP sites do not interfere with the wild-type pattern of gene expression. Current technology for homologous recombination in ES cells requires the concomitant insertion of selectable markers into the target gene. These selectable markers are expression cassettes conferring antibiotic resistance, such as the Neomycin resistance gene. These cassettes contain strong promoters and prokaryotic DNA and their presence can disturb the expression of the targeted and sometimes also surrounding genes in unpredictable ways (Meyers et al., 1998). To prevent the possibility of unwanted effects on target genes the removal of the selectable marker is recommended in gene knockout strategies.

For this reason, we decided to use tri-lox strategy. This strategy involves the use of three loxP sites, positioned such that neo, as well as the essential region of the gene of interest, is floxed. A partial Cre-mediated recombination event then removes neo, leaving the essential region floxed.

To circumvent the possible early embryonic lethality and to study the later functional significance of HBP1 in individual organs, a conditional knockout strategy involving the bacterial Cre/loxP recombinase system was chosen. A study comparing the recombination frequencies of loxP sites in different alleles in vivo found marked differences in recombination frequencies. The distance between LoxP sites in the

various targets tested explained in part these differences. The construct with the smallest floxed region showed a higher percentage of deletion of the floxed allele compared to the alleles with larger floxed regions (Vooijs et al., 2001). To keep the loxP sites as close as possible to each other to ensure a higher degree of recombination, we put loxP sites around a crucial region of the HBP1 gene rather then around the entire HBP1 locus.

Transcription factors are defined as proteins that bind to regulatory regions in the genome and help control gene expression. As sown previously in this thesis, removal of the HMG box allows the protein to stay in the cytoplasm and stops it from binding to DNA. Thus, removal of the HMG box will stop HBP1 functioning as a transcription factor. Unfortunately, during the course of this work, it was reported that HBP1-DNA binding is not required for inhibition of Wnt-ß-catenin transcriptional activation (Sampson et al., 2001). This means that removing the DNA binding domain will not abolish all the functions of HBP1. Thus, the results obtained only address the function of HBP1 as a transcription factor when its DNA binding activity is indispensable.

To flox the HMG box a floxed neo resistance marker, a necessary part of the initial floxed allele, was cloned 5' of the HMG box and a single loxP site was cloned 3' of the HMG box.

Regions containing multiple loxP sites can be partially eliminated by transient Cre-expression. As shown in figure 20, the HBP1 floxed allele contained three loxP sites. Cre mediated recombination could theoretically occur between any two of the three

loxP sites, generating offspring with three possible genotypes. Mice with different genotypes were identified by southern analysis. This strategy yielded the complete HMG box knockout and a strain devoid of neo but containing the floxed HMG box. Deletion of the HMG box alone (without deleting the neo) or no recombination was not observed in the animals screened.

Although based on no more than circumstantial evidence, it seems that partial deletions of tri-lox alleles are biased towards one of the two possible forms. The analysis of the presented tri-lox allele indicates that the neo is preferentially deleted compared to the HMG box. It has to be noted that the distance of loxP sites surrounding the neo was smaller then the distance of loxP sites surrounding the HMG box. However, at least for the ranges used here, it seems unlikely that this will influence the partial Cre lox deletion. Furthermore, other studies using the tri-lox strategy have also seen a preferential deletion of the neo gene (Holzenberger et al., 2000; Leneuve et al., 2003). It seems more likely that the proximity to prokaryotic DNA, the natural environment for loxP sites, the nucleotide sequences immediately surrounding the loxP sites, or the state of chromosomal structure may affect the likelihood of recombination.

In vivo and in vitro protocols have been established to remove the neo cassette from the targeted allele. In the first protocols the tri-lox configurations were converted into final floxed conditional alleles by the transient transfection of targeted ES cell clones in vitro with Cre-expressing plasmids, prior to chimera production(Gu et al., 1994). While this method has been used successfully in generating conditional knockouts, it does require additional manipulation of ES cells. This means additional

electroporation and prolonged culture of already targeted ES cells. This may reduce ES cell totipotency and the probability of germ-line chimeras developing from a particular clone.

To circumvent the extra ES manipulation it is possible to eliminate the floxed marker fragment selectively in vivo. The established protocols use Cre transgenic mouse lines capable of inducing partial loxP recombination (Holzenberger et al., 2000; Leneuve et al., 2003). Heterozygous F0 tri-lox mice are bred to the Cre transgenic mouse line to obtain double transgenics in the F1 generation, which should be mosaic for the Cre deletion. These mosaic mice are then crossed with wild-type mice. F2 animals are then screened for the desired recombination. Thus, an additional mouse generation (9–10 weeks) is required to accomplish selection cassette excision.

Here we explored an alternative strategy based on partial in vivo Cre recombination in the germline and allelic segregation in the next generation. We used ES cells that contain a Protamine Cre expression cassette for in vivo Cre expression. It was previously shown that these ES cells can mediate the efficient recombination of a Cre target transgene in the male germ line, but not in other tissues (O'Gorman et al., 1997).

This has the advantage that the recombination occurs in the germ-line of F0 (chimeras). F1 animals are then screened for the desired recombination. This saves the need for an F2 generation to establish the mouse line. Thus, this method saves 9 to 10 weeks in breeding time compared to the previous described method. Moreover, in the in vivo strategy, the initial ES cell clones, obtained by homologous recombination, are

available for blastocyst injection at an earlier time-point, without additional in vitro manipulations, so that unexpected difficulties with the germ-line transmission of particular clones or mutations will be detected sooner.

HBP1 knockout mice

To our surprise homozygous HBP1 knockout mice were born. This might be due to the fact that we only removed the DNA binding domain of HBP1. We have only abolished the function of HBP1 when the DNA binding is indispensable for its functioning. In all other cases, the HBP1 mutant without the HMG box might still be functioning as wildtype.

Although both HBP1$^{\Delta HMG/\Delta HMG}$ and HBP1 $^{flHMG/flHMG}$ were obtained in this thesis efforts were concentrated on the HBP1$^{\Delta HMG/\Delta HMG}$ mice. Genomic southern blot and PCR analyses proved the proper homologous recombination. Northern blot hybridizations were performed to control for the presence of the right-truncated mRNA transcripts. In HBP1$^{\Delta HMG/\Delta HMG}$ mice, the expected truncated HBP1 transcript lacking the 470 nucleotides, corresponding to exons 8 and 9, was found. The combined Southern and Northern analyses proved that the targeting event had removed the portion of the gene encoding the HMG box domain.

It remains possible that the residual expressed N-terminal region of HBP1, containing the AXH domain, retains all the functional properties of the full-length protein. To dismiss the possibility we tested whether the truncated protein still can act as a

transcription factor. As said before transcription factors are defined by their ability to bind DNA. Furthermore, it was shown previously that full length HBP1 interacts with the FT1 region of the LCR in a sequence-specific manner and contributes to the regulation of LCR function (Zhuma et al., 1999). Thus, if we could show that the binding to FT1 by HBP1 was abolished in the knockout mice, we could conclude that truncated HBP1 was not longer acting as a transcription factor. To demonstrate that the truncated HBP1 has lost the ability to bind DNA, a band-shift experiment was performed using the thymus nuclear extract with an oligo-nucleotide containing the HBP1 HMG box-binding motif. The experiment revealed that $HBP1^{\Delta HMG/\Delta HMG}$ nuclear extracts have lost the ability to bind to the HBP1 recognition sequences in electrophoretic mobility shift assays (figure 24). Thus, the $HBP1^{\Delta HMG/\Delta HMG}$ mice lack the ability to bind to the hCD2 LCR. This confirms that deletion of the HMG box inactivates HBP1 whenever DNA binding is indispensable. We have to assume that HBP1 act as wild type in all cases where DNA binding is not necessary for its function.

Functional analysis of T cells in $HBP1^{\Delta HMG/\Delta HMG}$ mice

We have demonstrated that HBP1 knockout mice are generally healthy and that they lack HBP1 function. These mice develop normally in utero, as the knockout allele is inherited in a ratio expected by Mendelian genetics. No clear phenotypic abnormalities were noted upon assessment of fresh tissues. Taken together, these findings indicate that the DNA binding activity and sub-cellular localisation of HBP1 DNA binding is dispensable for development and function through early adulthood.

K.M. Lin et al (Lin et al., 2001a) suggested that HBP1 is actively involved in the differentiation process during haematopoiesis. Furthermore, the haematopoietic system provides an excellent model for studying the function of HBP1 in cell proliferation, lineage commitment, differentiation, and maturation because each stage is well defined by the expression of cell surface molecules. Therefore, we decided to inspect the haematological compartment in greater detail. Preliminary FACS analysis of the T cell, B cell and myeloid lineages in homozygous and heterozygous HBP1 knockout mice revealed no obvious abnormalities when compared with wild-type littermates. These results are preliminary, and we are currently carrying out further detailed analyses on larger numbers of mice of both sexes and different ages to detect whether there are more subtle effects. As no alterations in steady-state haematopoiesis were observed, we investigated how T cells from these mice react to apoptotic and proliferative stimuli.

HBP1 and apoptosis

Results from HBP1 transfection experiments suggest that the HBP1 protein may induce apoptosis. Thus, one would predict that inactivation of the endogenous HBP1 should inhibit the apoptotic death process. This hypothesis was investigated by stimulating thymocytes from HBP1$^{\Delta HMG/\Delta HMG}$ mice with different apoptotic stimuli. Thymocytes from HBP1$^{\Delta HMG/\Delta HMG}$ mice exhibit roughly the same levels of apoptosis after treatment with low dose γ-radiation or dexamethasone. These data indicate that even though multiple apoptotic pathways were activated no effect was seen on the rate

of apoptosis in HBP1$^{\Delta HMG/\Delta HMG}$ thymocytes. These seemingly conflicting results might be due to a dosage effect of HBP1 on apoptosis and only when high levels of HBP1 are present an effect on apoptosis is seen. These results might also indicate that the results obtained with the EYFP fusion protein are an artificial artefact. More experiments need to be done to address these questions.

HBP1 and proliferation

HBP1 has been shown to be a negative regulator of the cell cycle. Accordingly, ablation of HBP1 might increase the proliferation potential of cells from these mice. To investigate the possible role of HBP1 in modulating mitogenic signalling of lymphocytes we assessed aspects of cellular proliferation of T cells of HBP1$^{\Delta HMG/\Delta HMG}$ and wild type mice subsequent to T-cell receptor stimulation. To follow the fate of individual T cells throughout activation and clonal expansion, we loaded the cells with the fluorescent dye carboxyfluorescein succinimidyl ester (CFSE) prior to stimulation. CFSE segregates equally between daughter cells upon cell division resulting in the sequential halving of cellular fluorescence intensity with each successive generation. When analyzed by flow cytometry, this sequential halving of fluorescence is visualized as reduced fluorescent intensity of cohorts of cells and, thus, can be used to track cell division in populations of proliferating cells. It was found that within 3 days of stimulation T cells from HBP1$^{\Delta HMG/\Delta HMG}$ mice respond more vigorously to the CD3 stimulation (as shown by increased CFSE dilution) compared to T cells from wild type littermates. This was true for both CD4 and CD8 T cells from HBP1$^{\Delta HMG/\Delta HMG}$ mice although the difference seen was larger for CD8 T

cells compared to CD4 T cells. CD8 cells are more responsive to anti-CD3 stimulation in vitro compared to CD4 cells. Thus, it is unclear if HBP1 differentially affects CD4 and CD8 cells or if the difference is due to the assay used.

HBP1 may control either entry into cell cycle from resting cells or progression through the cell cycle. Analysis of cell division showed that the fraction of cells triggered through the TCR to divide (percent divided) but not the mean number of divisions of the dividing cells (proliferation index) was significantly increased in both CD4 and CD8 T cells. Together this might indicate that HBP1 prevents the cell from entering the cell cycle but does not affect the cell cycle once the cells are in cycle.

It is possible that HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes respond with increased cell proliferation because they are in a "poised for activation" state and, thus, predisposed for immediate response to stimulation. Previously, gene-targeting of LKF and JAK3 (Grossman et al., 1999; Kuo et al., 1997) demonstrated the presence of such pre-activated lymphocytes with elevated CD69 and/or CD25 levels. To explore this possibility, FACS analysis of HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes for CD69 and CD25 was performed. CD69 and CD25 were absent from non-stimulated wild-type T cells, while intermediate levels of CD69 and CD25 expression were found on non-stimulated HBP1$^{\Delta HMG/\Delta HMG}$ lymphocytes. This indicates that these cells might be in a pre-activated state, even in non-stimulating conditions. Post-stimulation the differences in the amount of CD69 and CD25 expression were still observed. From these initial experiments, we conclude that T cells in HBP1$^{\Delta HMG/\Delta HMG}$ mice are in a pre-activated state and are hyper-responsive to signals through the TCR.

Possible roles for HBP1.

Our results with the HBP1$^{\Delta HMG/\Delta HMG}$ mice, in which the HMG box of HBP1 has been removed, support a role for HBP1 as a cell cycle inhibitor in differentiated tissues. HBP1 is constantly up regulated in differentiation and under conditions of cell cycle arrest in muscle, myeloid, erythroid, and other cell types while it is expressed at low levels in cycling cells. Consistent with this, we showed that lymphocytes from HBP1$^{\Delta HMG/\Delta HMG}$ mice are more likely to enter the cell cycle compared to wildtype mice while cycling cells underwent the same amount of mean divisions. Furthermore, over expression of HBP1 in transgenic animals resulted in a delayed G1 phase of the first cell cycle. The progression of G1 phase in the subsequent cell cycles was not determined (Shih et al., 2001). Thus, HBP1 blocks cell cycle entry but does not affect the cell cycle of cycling cells.

We propose that HBP1 sets the cellular threshold to mitogenic signals. In this model increased levels of HBP1 dampens the mitogenic response while cells with low levels of HBP1 have an increased mitogenic response. Consistent with this the lymphocytes lacking fully functional HBP1 are "poised for activation".

It is unclear at the moment how HBP1 regulates the threshold to mitogenic signals. HBP1 can form a complex with retinoblastoma and this complex can enforce a proliferation barrier in differentiated tissues (Tevosian et al., 1997). HBP1 alone or in complex with retinoblastoma can affect the expression of different target genes such as the cell cycle regulators cyclin D1 and N-myc and genes associated with terminal

differentiation such as histone $H1^0$. It is possible that the phenotype seen by HBP1 is through the regulation of expression of these target genes. Future experiments will address the question of target gene expression by a combination of western and northern blot analysis.

Clinical relevance

A major question for tumour suppression is how normal proliferation controls are lost in the early stages of certain cancers, in which there is a proliferation of differentiated cells. The phenotype observed in lymphocytes from $HBP1^{\Delta HMG/\Delta HMG}$ mice is clearly not leukaemia, the hyper-responsiveness of the cells indicates that certain pathways that regulate the cell cycle are affected and mis-regulation of the cell cycle might lead to cancer. The hyper-proliferative response in $HBP1^{\Delta HMG/\Delta HMG}$ mice suggests that HBP1 could have a tumour suppressor function. Consistent with the notion of tumour suppression, the HBP1 gene resides in human chromosome 7q31. Loss of chromosome 7 or deletions of its long arm (7q-) are recurring chromosome abnormalities in myeloid disorders. These aberrations are associated with myelodysplastic syndrome (MDS) and acute myeloid leukaemia (AML), in particular with therapy-related MDS/AML (t-MDS/t-AML) after therapy with alkylating agents or secondary MDS/AML after occupational exposure to chemical mutagens. A commonly deleted region was defined by the MDS/AML cases and was a genomic region of approximately 20 Mb encompassing the terminal part of band 7q22 and the entire band 7q31. Although the commonly deleted genomic region is still large, HBP1 is a good candidate as a tumour-suppressing gene.

Thus, a loss-of-function mutation in HBP1 results in an enhanced cell cycle response in the lymphocytes and may be important in tumour onset. Future work will be directed at the importance of HBP1 in tumour suppression and in signalling networks that govern proliferation.

References

Ahringer, J. (2000). NuRD and SIN3 histone deacetylase complexes in development. Trends Genet *16*, 351-356.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I. L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193-197.

Allshire, R. C., Javerzat, J. P., Redhead, N. J., and Cranston, G. (1994). Position effect variegation at fission yeast centromeres. Cell *76*, 157-169.

Ambrosino, C., and Nebreda, A. R. (2001). Cell cycle regulation by p38 MAP kinases. Biol Cell *93*, 47-51.

Austin, G. E., and Lin, K. W. M. (2002). Regulation in early myeloid cells of the expression of HBP1, a transcription factor which controls DNA proliferation in normal and neoplastic cells. Modern Pathology *15*, 965.

Becker, P. B., and Horz, W. (2002). ATP-dependent nucleosome remodeling. Annu Rev Biochem *71*, 247-273.

Bell, A. C., and Felsenfeld, G. (1999). Stopped at the border: boundaries and insulators. Curr Opin Genet Dev *9*, 191-198.

Bewley, C. A., Gronenborn, A. M., and Clore, G. M. (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. Annu Rev Biophys Biomol Struct *27*, 105-131.

Bierer, B. E., Peterson, A., Gorga, J. C., Herrmann, S. H., and Burakoff, S. J. (1988). Synergistic T cell activation via the physiological ligands for CD2 and the T cell receptor. J Exp Med *168*, 1145-1156.

Bird, A. P., and Wolffe, A. P. (1999). Methylation-induced repression--belts, braces, and chromatin. Cell *99*, 451-454.

Blackwood, E. M., and Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. Science *281*, 61-63.

Blau, H. M., Brazelton, T. R., and Weimann, J. M. (2001). The evolving concept of a stem cell: entity or function? Cell *105*, 829-841.

Brantjes, H., Barker, N., van Es, J., and Clevers, H. (2002). TCF: Lady Justice casting the final verdict on the outcome of Wnt signalling. Biol Chem *383*, 255-261.

Burright, E. N., Davidson, J. D., Duvick, L. A., Koshy, B., Zoghbi, H. Y., and Orr, H. T. (1997). Identification of a self-association region within the SCA1 gene product, ataxin-1. Hum Mol Genet *6*, 513-518.

Busslinger, M., Nutt, S. L., and Rolink, A. G. (2000). Lineage commitment in lymphopoiesis. Curr Opin Immunol *12*, 151-158.

Butler, J. E., and Kadonaga, J. T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev *16*, 2583-2592.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat Genet *32*, 623-626.

Chen, S. L., Loffler, K. A., Chen, D., Stallcup, M. R., and Muscat, G. E. (2002). The coactivator-associated arginine methyltransferase is necessary for muscle differentiation: CARM1 coactivates myocyte enhancer factor-2. J Biol Chem *277*, 4324-4333.

Cheung, P., Tanner, K. G., Cheung, W. L., Sassone-Corsi, P., Denu, J. M., and Allis, C. D. (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell *5*, 905-915.

Clark, A. J., Harold, G., and Yull, F. E. (1997). Mammalian cDNA and prokaryotic reporter sequences silence adjacent transgenes in transgenic mice. Nucleic Acids Res *25*, 1009-1014.

Cook, S. J., Balmanno, K., Garner, A., Millar, T., Taverner, C., and Todd, D. (2000). Regulation of cell cycle re-entry by growth, survival and stress signalling pathways. Biochem Soc Trans *28*, 233-240.

Coqueret, O. (2002). Linking cyclins to transcriptional control. Gene *299*, 35-55.

Cui, C., Wani, M. A., Wight, D., Kopchick, J., and Stambrook, P. J. (1994). Reporter genes in transgenic mice. Transgenic Res *3*, 182-194.

de Boer, C. J., van Krieken, J. H., Schuuring, E., and Kluin, P. M. (1997). Bcl-1/cyclin D1 in malignant lymphoma. Ann Oncol *8 Suppl 2*, 109-117.

de Boer, J., Williams, A., Skavdis, G., Harker, N., Coles, M., Tolaini, M., Norton, T., Williams, K., Roderick, K., Potocnik, A. J., and Kioussis, D. (2003). Transgenic mice with hematopoietic and lymphoid specific expression of Cre. Eur J Immunol *33*, 314-325.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Doenecke, D., Albig, W., Bode, C., Drabent, B., Franke, K., Gavenis, K., and Witt, O. (1997). Histones: genetic diversity and tissue-specific gene expression. Histochem Cell Biol *107*, 1-10.

Dyson, N. (1994). pRB, p107 and the regulation of the E2F transcription factor. J Cell Sci Suppl *18*, 81-87.

Eberharter, A., and Becker, P. B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. EMBO Rep *3*, 224-229.

Fernandez-Funez, P., Nino-Rosales, M. L., de Gouyon, B., She, W. C., Luchak, J. M., Martinez, P., Turiegano, E., Benito, J., Capovilla, M., Skinner, P. J., *et al.* (2000). Identification of genes that modify ataxin-1-induced neurodegeneration. Nature *408*, 101-106.

Festenstein, R., Sharghi-Namini, S., Fox, M., Roderick, K., Tolaini, M., Norton, T., Saveliev, A., Kioussis, D., and Singh, P. (1999). Heterochromatin protein 1 modifies mammalian PEV in a dose- and chromosomal-context-dependent manner. Nat Genet *23*, 457-461.

Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M., and Kioussis, D. (1996). Locus control region function and heterochromatin-induced position effect variegation. Science *271*, 1123-1125.

Fry, C. J., and Peterson, C. L. (2002). Transcription. Unlocking the gates to gene expression. Science *295*, 1847-1848.

Gartel, A. L., Goufman, E., Tevosian, S. G., Shih, H., Yee, A. S., and Tyner, A. L. (1998). Activation and repression of p21(WAF1/CIP1) transcription by RB binding proteins. Oncogene *17*, 3463-3469.

Georgopoulos, K., Moore, D. D., and Derfler, B. (1992). Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. Science *258*, 808-812.

Giles, R. H., van Es, J. H., and Clevers, H. (2003). Caught up in a Wnt storm: Wnt signaling in cancer. Biochim Biophys Acta *1653*, 1-24.

Good, L. (2003). Translation repression by antisense sequences. Cell Mol Life Sci *60*, 854-861.

Goodwin, G. H., and Johns, E. W. (1973). Isolation and characterisation of two calf-thymus chromatin non-histone proteins with high contents of acidic and basic amino acids. Eur J Biochem *40*, 215-219.

Grandori, C., Cowley, S. M., James, L. P., and Eisenman, R. N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu Rev Cell Dev Biol *16*, 653-699.

Greaves, D. R., Wilson, F. D., Lang, G., and Kioussis, D. (1989). Human CD2 3'-flanking sequences confer high-level, T cell-specific, position-independent gene expression in transgenic mice. Cell *56*, 979-986.

Grossman, W. J., Verbsky, J. W., Yang, L., Berg, L. J., Fields, L. E., Chaplin, D. D., and Ratner, L. (1999). Dysregulated myelopoiesis in mice lacking Jak3. Blood *94*, 932-939.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell *51*, 975-985.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science *265*, 103-106.

Harbour, J. W., and Dean, D. C. (2000). The Rb/E2F pathway: expanding roles and emerging paradigms. Genes Dev *14*, 2393-2409.

Harper, J. W., and Elledge, S. J. (1996). Cdk inhibitors in development and cancer. Curr Opin Genet Dev *6*, 56-64.

Henikoff, S. (1990). Position-effect variegation after 60 years. Trends Genet *6*, 422-426.

Holzenberger, M., Lenzner, C., Leneuve, P., Zaoui, R., Hamard, G., Vaulont, S., and Bouc, Y. L. (2000). Cre-mediated germline mosaicism: a method allowing rapid generation of several alleles of a target gene. Nucleic Acids Res *28*, E92.

Horn, P. J., and Peterson, C. L. (2001). The bromodomain: a regulator of ATP-dependent chromatin remodeling? Front Biosci *6*, D1019-1023.

Jeanmougin, F., Wurtz, J. M., Le Douarin, B., Chambon, P., and Losson, R. (1997). The bromodomain revisited. Trends Biochem Sci *22*, 151-153.

Kioussis, D., and Festenstein, R. (1997). Locus control regions: overcoming heterochromatin-induced gene inactivation in mammals. Curr Opin Genet Dev *7*, 614-619.

Kleinjan, D. J., and van Heyningen, V. (1998). Position effect in human genetic disease. Hum Mol Genet *7*, 1611-1618.

Knudson, A. G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A *68*, 820-823.

Kouzarides, T. (2002). Histone methylation in transcriptional control. Curr Opin Genet Dev *12*, 198-209.

Kuo, C. T., Veselits, M. L., and Leiden, J. M. (1997). LKLF: A transcriptional regulator of single-positive T cell quiescence and survival. Science *277*, 1986-1990.

Lachner, M., and Jenuwein, T. (2002). The many faces of histone lysine methylation. Curr Opin Cell Biol *14*, 286-298.

Lake, R. A., Wotton, D., and Owen, M. J. (1990). A 3' transcriptional enhancer regulates tissue-specific expression of the human CD2 gene. Embo J *9*, 3129-3136.

Lang, G., Mamalaki, C., Greenberg, D., Yannoutsos, N., and Kioussis, D. (1991). Deletion analysis of the human CD2 gene locus control region in transgenic mice. Nucleic Acids Res *19*, 5851-5856.

Lang, G., Wotton, D., Owen, M. J., Sewell, W. A., Brown, M. H., Mason, D. Y., Crumpton, M. J., and Kioussis, D. (1988). The structure of the human CD2 gene and its expression in transgenic mice. Embo J *7*, 1675-1682.

Latchman, D. S. (1997). Transcription factors: an overview. Int J Biochem Cell Biol *29*, 1305-1312.

Lavender, P., Vandel, L., Bannister, A. J., and Kouzarides, T. (1997). The HMG-box transcription factor HBP1 is targeted by the pocket proteins and E1A. Oncogene *14*, 2721-2728.

Lee, M. H., and Yang, H. Y. (2001). Negative regulators of cyclin-dependent kinases and their roles in cancers. Cell Mol Life Sci *58*, 1907-1922.

Lemercier, C., Duncliffe, K., Boibessot, I., Zhang, H., Verdel, A., Angelov, D., and Khochbin, S. (2000). Involvement of retinoblastoma protein and HBP1 in histone H1(0) gene expression. Mol Cell Biol *20*, 6627-6637.

Leneuve, P., Colnot, S., Hamard, G., Francis, F., Niwa-Kawakita, M., Giovannini, M., and Holzenberger, M. (2003). Cre-mediated germline mosaicism: a new transgenic mouse for the selective removal of residual markers from tri-lox conditional alleles. Nucleic Acids Res *31*, e21.

Li, Q., Peterson, K. R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood *100*, 3077-3086.

Lin, K. M., Zhao, W. G., Bhatnagar, J., Zhao, W. D., Lu, J. P., Simko, S., Schueneman, A., and Austin, G. E. (2001a). Cloning and expression of human HBP1, a high mobility group protein that enhances myeloperoxidase (MPO) promoter activity. Leukemia *15*, 601-612.

Lin, K. W. M., Bhatnagar, J., Murphy, M. L., and Austin, G. E. (2001b). Regulation of transcription of the human HBP1 gene during myeloid differentiation. Blood *98*, 4172.

Lipinski, M. M., and Jacks, T. (1999). The retinoblastoma gene family in differentiation and development. Oncogene *18*, 7873-7882.

Meyers, E. N., Lewandoski, M., and Martin, G. R. (1998). An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. Nat Genet *18*, 136-141.

Mirro, J., Jr. (1992). Pathology and immunology of acute leukemia. Leukemia *6 Suppl 4*, 13-15.

Morales, V., and Richard-Foy, H. (2000). Role of histone N-terminal tails and their acetylation in nucleosome dynamics. Mol Cell Biol *20*, 7230-7237.

Morgan, D. O. (1995). Principles of CDK regulation. Nature *374*, 131-134.

Mulligan, G., and Jacks, T. (1998). The retinoblastoma gene family: cousins with overlapping interests. Trends Genet *14*, 223-229.

Mushegian, A. R., Bassett, D. E., Jr., Boguski, M. S., Bork, P., and Koonin, E. V. (1997). Positionally cloned human disease genes: patterns of evolutionary conservation and functional motifs. Proc Natl Acad Sci U S A *94*, 5831-5836.

Nagy, A., Moens, C., Ivanyi, E., Pawling, J., Gertsenstein, M., Hadjantonakis, A. K., Pirity, M., and Rossant, J. (1998). Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles. Curr Biol *8*, 661-664.

Nakayama, K. (1998). Cip/Kip cyclin-dependent kinase inhibitors: brakes of the cell cycle engine during development. Bioessays *20*, 1020-1029.

Narlikar, G. J., Fan, H. Y., and Kingston, R. E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Nicoletti, I., Migliorati, G., Pagliacci, M. C., Grignani, F., and Riccardi, C. (1991). A rapid and simple method for measuring thymocyte apoptosis by propidium iodide staining and flow cytometry. J Immunol Methods *139*, 271-279.

Novak, A., and Dedhar, S. (1999). Signaling through beta-catenin and Lef/Tcf. Cell Mol Life Sci *56*, 523-537.

Ogilvy, S., Metcalf, D., Gibson, L., Bath, M. L., Harris, A. W., and Adams, J. M. (1999). Promoter elements of vav drive transgene expression in vivo throughout the hematopoietic compartment. Blood *94*, 1855-1863.

O'Gorman, S., Dagenais, N. A., Qian, M., and Marchuk, Y. (1997). Protamine-Cre recombinase transgenes efficiently recombine target sequences in the male germ line of mice, but not in embryonic stem cells. Proc Natl Acad Sci U S A *94*, 14602-14607.

Osoegawa, K., Tateno, M., Woon, P. Y., Frengen, E., Mammoser, A. G., Catanese, J. J., Hayashizaki, Y., and de Jong, P. J. (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res *10*, 116-128.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Rossant, J., and Nagy, A. (1995). Genome engineering: the new mouse genetics. Nat Med *1*, 592-594.

Roth, S. Y., Denu, J. M., and Allis, C. D. (2001). Histone acetyltransferases. Annu Rev Biochem *70*, 81-120.

Sampson, E. M., Haque, Z. K., Ku, M. C., Tevosian, S. G., Albanese, C., Pestell, R. G., Paulson, K. E., and Yee, A. S. (2001). Negative regulation of the Wnt-beta-catenin pathway by the transcriptional repressor HBP1. Embo J *20*, 4500-4511.

Sasada, T., and Reinherz, E. L. (2001). A critical role for CD2 in both thymic selection events and mature T cell function. J Immunol *166*, 2394-2403.

Shekhar, M. P., Lyakhovich, A., Visscher, D. W., Heng, H., and Kondrat, N. (2002). Rad6 overexpression induces multinucleation, centrosome amplification, abnormal mitosis, aneuploidy, and transformation. Cancer Res *62*, 2115-2124.

Shih, H. H., Tevosian, S. G., and Yee, A. S. (1998). Regulation of differentiation by HBP1, a target of the retinoblastoma protein. Mol Cell Biol *18*, 4732-4743.

Shih, H. H., Xiu, M., Berasi, S. P., Sampson, E. M., Leiter, A., Paulson, K. E., and Yee, A. S. (2001). HMG box transcriptional repressor HBP1 maintains a proliferation barrier in differentiated liver tissue. Mol Cell Biol *21*, 5723-5732.

Shimshek, D. R., Kim, J., Hubner, M. R., Spergel, D. J., Buchholz, F., Casanova, E., Stewart, A. F., Seeburg, P. H., and Sprengel, R. (2002). Codon-improved Cre recombinase (iCre) expression in the mouse. Genesis *32*, 19-26.

Soullier, S., Jay, P., Poulat, F., Vanacker, J. M., Berta, P., and Laudet, V. (1999). Diversification pattern of the HMG and SOX family members during evolution. J Mol Evol *48*, 517-527.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol *98*, 503-517.

Springer, T. A. (1990). Adhesion receptors of the immune system. Nature *346*, 425-434.

Srinivas, S., Watanabe, T., Lin, C. S., William, C. M., Tanabe, Y., Jessell, T. M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev Biol *1*, 4.

Stauber, R., Gaitanaris, G. A., and Pavlakis, G. N. (1995). Analysis of trafficking of Rev and transdominant Rev proteins in living cells using green fluorescent protein fusions: transdominant Rev blocks the export of Rev from the nucleus to the cytoplasm. Virology *213*, 439-449.

Steff, A. M., Fortin, M., Arguin, C., and Hugo, P. (2001). Detection of a decrease in green fluorescent protein fluorescence for the monitoring of cell death: an assay amenable to high-throughput screening technologies. Cytometry *45*, 237-243.

Strebel, A., Harr, T., Bachmann, F., Wernli, M., and Erb, P. (2001). Green fluorescent protein as a novel tool to measure apoptosis and necrosis. Cytometry *43*, 126-133.

Sudo, T., Nishikawa, S., Ohno, N., Akiyama, N., Tamakoshi, M., and Yoshida, H. (1993). Expression and function of the interleukin 7 receptor in murine lymphocytes. Proc Natl Acad Sci U S A *90*, 9125-9129.

Sun, Z. W., and Allis, C. D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104-108.

Swanson, K. A., Knoepfler, P. S., Huang, K., Kang, R. S., Cowley, S. M., Laherty, C. D., Eisenman, R. N., and Radhakrishnan, I. (2004). HBP1 and Mad1 repressors bind the Sin3 corepressor PAH2 domain with opposite helical orientations. Nat Struct Mol Biol *11*, 738-746.

Tevosian, S. G., Shih, H. H., Mendelson, K. G., Sheppard, K. A., Paulson, K. E., and Yee, A. S. (1997). HBP1: a HMG box transcriptional repressor that is targeted by the retinoblastoma family. Genes Dev *11*, 383-396.

Thomas, K. R., and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Thyagarajan, B., Guimaraes, M. J., Groth, A. C., and Calos, M. P. (2000). Mammalian genomes contain active recombinase recognition sites. Gene *244*, 47-54.

Tybulewicz, V. L., Ardouin, L., Prisco, A., and Reynolds, L. F. (2003). Vav1: a key signal transducer downstream of the TCR. Immunol Rev *192*, 42-52.

van der Merwe, P. A., Barclay, A. N., Mason, D. W., Davies, E. A., Morgan, B. P., Tone, M., Krishnam, A. K., Ianelli, C., and Davis, S. J. (1994). Human cell-adhesion molecule CD2 binds CD58 (LFA-3) with a very low affinity and an extremely fast dissociation rate but does not bind CD48 or CD59. Biochemistry *33*, 10149-10160.

Varga-Weisz, P. (2001). ATP-dependent chromatin remodeling factors: nucleosome shufflers with many missions. Oncogene *20*, 3076-3085.

Vooijs, M., Jonkers, J., and Berns, A. (2001). A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent. EMBO Rep *2*, 292-297.

Wallrath, L. L., and Elgin, S. C. (1995). Position effect variegation in Drosophila is associated with an altered chromatin structure. Genes Dev *9*, 1263-1277.

Weinberg, R. A. (1991). Tumor suppressor genes. Science *254*, 1138-1146.

Weiss, M. A. (2001). Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition. Mol Endocrinol *15*, 353-362.

Wilson, T. J., Cowdery, H. E., Xu, D., Kola, I., and Hertzog, P. J. (2002). A human CD2 minigene directs CRE-mediated recombination in T cells in vivo. Genesis *33*, 181-184.

Xiu, M., Kim, J., Sampson, E., Huang, C. Y., Davis, R. J., Paulson, K. E., and Yee, A. S. (2003). The transcriptional repressor HBP1 is a target of the p38 mitogen-activated protein kinase pathway in cell cycle regulation. Mol Cell Biol *23*, 8890-8901.

Yadav, N., Lee, J., Kim, J., Shen, J., Hu, M. C., Aldaz, C. M., and Bedford, M. T. (2003). Specific protein methylation defects and gene expression perturbations in coactivator-associated arginine methyltransferase 1-deficient mice. Proc Natl Acad Sci U S A *100*, 6464-6468.

Yee, A. S., Paulson, E. K., McDevitt, M. A., Rieger-Christ, K., Summerhayes, I., Berasi, S. P., Kim, J., Huang, C. Y., and Zhang, X. (2004). The HBP1 transcriptional repressor and the p38 MAP kinase: unlikely partners in G1 regulation and tumor suppression. Gene *336*, 1-13.

Yelin, R., Dahary, D., Sorek, R., Levanon, E. Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., *et al.* (2003). Widespread occurrence of antisense transcription in the human genome. Nat Biotechnol *21*, 379-386.

Yue, S., Serra, H. G., Zoghbi, H. Y., and Orr, H. T. (2001). The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. Hum Mol Genet *10*, 25-30.

Zeng, L., and Zhou, M. M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS Lett *513*, 124-128.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.

Zhang, Z., Carriero, N., and Gerstein, M. (2004). Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends Genet *20*, 62-67.

Zheng, C., and Hayes, J. J. (2003). Structures and interactions of the core histone tail domains. Biopolymers *68*, 539-546.

Zhuma, T., Tyrrell, R., Sekkali, B., Skavdis, G., Saveliev, A., Tolaini, M., Roderick, K., Norton, T., Smerdon, S., Sedgwick, S., *et al.* (1999). Human HMG box transcription factor HBP1: a role in hCD2 LCR function. Embo J *18*, 6396-6406.

Zhumabekov, T., Corbella, P., Tolaini, M., and Kioussis, D. (1995). Improved version of a human CD2 minigene based vector for T cell-specific expression in transgenic mice. J Immunol Methods *185*, 133-140.